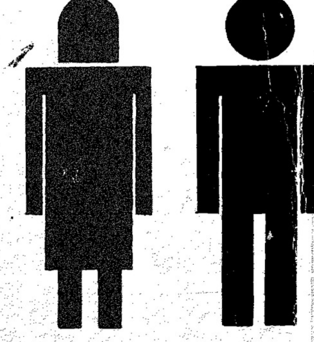
مقدمة في الإحصاء

الجزء الأول



الدكتور السيد نور
جامعة الإمارات العربية المتحدة






مقدمة في الإحصاء

بسم الله الرحمن الرحيم

مقدمة في الإحصاء

الجزء الأول

الدكتور
السيد نور
جامعة الإمارات العربية المتحدة



الطبعة الأولى
١٤٠٧ هـ - ١٩٨٧ م



دبي - طريق النفق - بناية الشيخ راشد القديمة .
هاتف ٤٣٣٨٨٦ - ص ب ١١٨١٧ - برقياً: قلمكم - تلكس ٤٦٩٦١ قلم إم

المحتويات

تصدير وشكر

هذا كتاب يشرح كيفية معالجة وتفسير البيانات الإحصائية ، وقد تم إعداده بهدف تنمية وعي القارىء العادي وقدراته على التعامل مع البيانات الرقمية ، واستنتاج الحقائق الأساسية فيها ، ثم التعبير عن هذه الحقائق وتوصيلها إلى الآخرين . ويهتم الكتاب بصفة أساسية بشرح الأسباب والمبررات وراء استخدام الأساليب الإحصائية المختلفة ، ولا يكتفي بمجرد توضيح الجانب الحسابي في هذه الأساليب . وبذلك يمكن للقارىء التعرف بشكل ناقد على أساليب التحليل الإحصائي واكتساب الخبرة في الحكم على مدى نجاح الآخرين في استخدام هذه الأساليب في مجالاتهم التطبيقية المختلفة .

ويعتبر الاعتماد على أمثلة ذات بيانات واقعية إحدى الخصائص المتميزة لهذا الكتاب . كذلك فإن أسلوب عرض الموضوعات المختلفة ومناقشتها لا يتطلب أن يكون قارىء هذا الكتاب على معرفة عميقة بالرياضيات ، إذ أن مستوى الرياضيات المفترض يقع في متناول الجميع .

وينقسم هذا الكتاب إلى جزءين ، الجزء الأول ويهتم بموضوعات الإحصاء الوصفي والجزء الثاني ويعالج أسس الاستنتاج الإحصائي . ويحتوي هذا الجزء الأول على عشرة أبواب ينتهي كل منها بمجموعة غزيرة من التمرينات العملية التي تعتمد على بيانات تنشأ في مجالات متعددة من مجالات التطبيق الإحصائي . ويعتبر حل هذه التمرينات متطلباً أساسياً لتحقيق مستوى جيد لفهم محتويات هذا الكتاب .

ولقد تم العمل على هذا الكتاب خلال عمل الكاتب بجامعة الإمارات العربية المتحدة ، وبذلك يمكن اعتبار الكتاب إحدى ثمرات فترة خدمته بهذه الجامعة . وما كان لعمل كهذا أن يتم دون تشجيع مِن الطلبة والزملاء . وهنا يسعدني الإشادة بفضل الكثيرين وأخص منهم كـل من الدكتـور مجدي سيـد مصطفى والدكتور عادل زاهر من أساتـذة الإحصاء بجـامعة الإمـارات اللذان حثـاني على اتمام هـذا الكتاب وقـامـا بمـراجعتـه والتعقيب عليـه عـدداً من المرات . فلهما شكري وامتناني .

السيد نور

تعريف بعلم الإحصاء

١ - مقدمة

تستخدم كلمة الإحصاء Statistics في أكثر من معنى . فقد جرت العادة على استخدام هذه الكلمة في معنى أول للدلالة على بيان رقمي يصف ظاهرة ما تتعلق بمفردة أو أكثر من مفردات مجتمع معين. وعلى ذلك فإن الإحصاءات الحيوية Vital Statistics تدل على البيانات الخاصة بأعداد المواليد والوفيات وحالات الزواج وحالات الطلاق ، كما أن الإحصاءات الاقتصادية Economic Statistics تشمل البيانات المتعلقة بقوة العمل والإنتاج والأسعار والمبيعات ، بينما تشير الإحصاءات الاجتماعية Social Statistics إلى بيانات السكان والإسكان والتعليم والجرائم والضمان الاجتماعي ، وهكذا .

وتستخدم كلمة الإحصاء في معنى ثان لوصف فرع من فروع العلم والمعرفة ، وهو المعنى الذي يمثل موضوع هذا الكتاب . وفي هذا الصدد فإن الإحصاء كعلم هو مجموعة الأساليب والمفاهيم التي تبحث في كيفية جمع وعرض وتحليل البيانات المتعلقة بظاهرة ما ، ثم طرق الاستفادة من هذه البيانات . ويمثل علم الاحصاء تبعاً لذلك جزءاً أساسياً من الأسلوب العلمي لدراسة المشكلات . ذلك أن الأسلوب العلمي يهدف ، كما هو معروف ، إلى توسيع آفاق المعرفة الإنسانية من خلال عملية مستمرة ومنظمة لاختبار الفروض والنظريات المختلفة بناءاً على بيانات ومعلومات موضوعية . وتختلف عناصر هذا الأسلوب من مجال علمي لآخر ، إلا أنه يمكن أن نحدد بشكل

عام أربع مراحل أساسية تشترك فيها معظم مجالات البحث العلمي وهي :

أ - **تحديد واضح لأهداف البحث** : قد يكون الهدف من البحث هو محاولة إثبات صحة نظرية أو فرض جديد ، أو محاولة التأكد من صحة نظرية أو فرض قائم بناءاً على البيانات المتاحة ، وقد يكون الهدف هو مجرد جمع معلومات تستخدم لوصف الوضع القائم في ظاهرة ما ، وقد يتعدى الهدف ذلك إلى محاولة التعرف على العوامل المختلفة التي تؤثر في ظاهرة ما بهدف الإستفادة من ذلك في بناء سياسات للتحكم في بعض جوانب تلك الظاهرة . ويتعين في جميع هذه الأحوال أن يكون هناك تحديداً واضحاً للهدف من الدراسة ، وذلك ضماناً لتحقيق الاستخدام الأمثل للإمكانات المتاحة وجمع البيانات المطلوبة لتحقيق هذا الهدف .

ب - **جمع البيانات** : ويعد ذلك أمراً حيوياً بالنسبة لأي بحث علمي . ويتم جمع البيانات ، في ضوء أهداف الدراسة ، من خلال مصادر عديدة تتراوح بين التجارب العلمية والمحاولات الميدانية والمسوح الاجتماعية والاقتصادية والسجلات المتاحة .

حـ - **تحليل البيانات** : ويقصد بذلك فحص البيانات بعناية وكيفية استخراج المعلومات المتوافرة فيها واستخدامها في الإجابة عن الأسئلة والقضايا التي تحددت كأهداف للبحث .

ء - **تحديد النتائج واتخاذ القرارات** : ويتم في هذه المرحلة تقويم نتائج تحليل البيانات في ضوء المعرفة المتاحة عن سلوك الظاهرة محل الدراسة ، وتحديد إمكانيات تعميم هذه النتائج وإبراز ما هو جديد ومبتكر فيها . وغالباً ما يؤدي ذلك إلى إثارة عدد من الأسئلة والقضايا الجديدة تتطلب دراستها البدء في إجراء بحوث أخرى ، ذلك أن البحث العلمي عملية مستمرة تتم في دورات متصلة تؤدي إحداها إلى الأخرى .

وقد ترتب على هذه العلاقة الوثيقة بين الإحصاء والأسلوب العلمي

للدراسة أن أصبحت المفاهيم الإحصائية أداة هامةً من أدوات البحث في الكثير من مجالات العلم والمعرفة مثل علوم الاجتماع والاقتصاد وادارة الأعمال والزراعة وعلوم الحياة والتربية والعلوم الهندسية والطب والقانون وغيرها . وقد ساعد على ذلك ما يتميز به العصر الحالي من توافر كم هائل من المعلومات والبيانات الرقمية عن كافة أوجه النشاط الإنساني ، وانتشار الحاسبات الآلية التي تساعد على سهولة التعامل مع هذه البيانات ، وما صاحب ذلك من تطور سريع في الأساليب الإحصائية اللازمة لتحليل هذه البيانات والاعتماد عليها كأساس لاتخاذ القرارات .

وتجدر الاشارة ، في هذا الصدد ، إلى أن إحدى المجلات العلمية المتخصصة قد قامت مؤخراً بالإشارة إلى الأساليب الإحصائية لدراسة المشكلات كواحدة من أهم عشرين اكتشافاً تمت خلال هذا القرن في مجالات العلم والتكنولوجيا والطب ، شأنها في ذلك شأن نظرية النسبية لأينشتاين وشأن اختراع البلاستيك . وقد جاء هذا الاختيار نتيجة لما ترتب على استخدام الأساليب الإحصائية من تغيير جذري في طرق البحث العلمي وأساليب اتخاذ القرارات .

٢ ـ أمثلة تطبيقية

نورد هنا بعض الأمثلة العملية في مجالات تطبيقية مختلفة توضح أهمية الطرق الإحصائية عند إجراء البحوث في إطار الأسلوب العلمي لدراسة المشكلات .

مثال (١) : قياس اتجاهات الرأي العام: قد يراد استطلاع آراء سكان المدينة حول نظام مقترح للمواصلات العامة ، أو قد يراد قياس مدى ارتياح طلبة الجامعة للسياسات التي ينتهجها مجلس اتحاد الطلبة ، أو قد يراد تحديد اتجاهات السكان حول قانون جديد للأحوال الشخصية ، أو قد يراد دراسة آراء المستهلكين في سلعة غذائية جديدة . . . الخ . يتم ذلك بجمع بيانات مناسبة ثم تحليلها واستخراج النتائج منها . وفي هذا الصدد يكون على الباحث أن

يحدد إجابات مقبولة للأسئلة الآتية :

(أ) ما هو عدد الأفراد الذين تجمع منهم البيانات ؟ وكيف يمكن اختيار هؤلاء الأفراد من المجتمع ؟

(ب) ما هو الموعد المناسب لجمع البيانات ؟

(حـ) ما هي نوع البيانات المطلوبة ؟ وما هي أساليب الحصول على هذه المعلومات من الأفراد ؟

(ء) هل البيانات دقيقة ؟ ما هي إمكانية تغيير الأفراد لآرائهم بعد الانتهاء من البحث ؟

(هـ) كيف سيتم تنظيم وتحليل البيانات التي تجمع ؟ وما هي إمكانيات تعميم نتائج البحث ؟ .

مثال (٢) : اختبار كفاءة مؤثر جديد أو المقارنة بين كفاءة عدد من المؤثرات . يكون الهدف من إجراء الدراسة في هذه الحالة هو الإجابة عن أسئلة من النوع التالي :

- هل يؤدي استخدام فيتامين (جـ) إلى الوقاية من الإصابة من البرد ؟
- هل يتميز النوع الجديد من بذور القمح عن غيره من الأنواع الأخرى ؟
- هل أدى برنامج تدريب المديرين إلى رفع كفاءتهم الإدارية ؟
- هل تعتبر الطريقة الجديدة لتدريس اللغة الانجليزية أفضل من غيرها ؟
- هل تتفق خصائص السلعة المنتجة في مصنع ما مع المواصفات المطلوبة ؟ .

ويتطلب الأمر ، في مثل هذه المواقف ، جمع البيانات اللازمة ثم تحليلها واتخاذ القرارات المناسبة . وتجدر الإشارة إلى ضرورة تصميم عملية جمع وتحليل البيانات في هذه الحالة بشكل يسمح بعزل تأثير العوامل الخارجية التي قد تؤثر في عملية المقارنة . فمثلاً ، إذا كان الهدف هو المقارنة بين كفاءة عدة أنواع من بذور القمح فإنه لا بد من التأكد من عزل تأثير عوامل التربة وظروف الطقس وطريقة الري ونوع السماد المستخدم على نتائج المقارنة .

مثال (٣) : التعرف على الوضع القائم في الظواهر الاقتصادية والاجتماعية المختلفة ، ثم التنبؤ بأنماط هذه الظواهر في المستقبل .
مثال ذلك ما يلي :

- ما هي خصائص التوزيع العمري والنوعي لسكان الدولة ؟
- ما هي أنماط الانفاق والاستهلاك السائدة بين أفراد المجتمع ؟ وما هي العوامل المختلفة التي لها علاقة بهذه الأنماط ؟
- ما هو معدل البطالة في المجتمع ؟ ما هو معدل مساهمة النساء في قوة العمل ؟
- ما هي أعداد الطلبة في مراحل النظام التعليمي المختلفة ؟
- ما هي أنماط النمو السكاني المتوقعة في الدولة حتى عام ٢٠٠٠؟
- ما هو عدد طلبة الجامعة المتوقع في عام ٢٠٠٠ ؟ وما هي التخصصات المتوقعة لهؤلاء الطلبة ؟
- ما هي أنماط الجرائم السائدة في المجتمع ؟ وكيف يمكن تقدير احتياجات مناطق الدولة المختلفة من رجال الشرطة ؟

ويتم في هذه الحالات جمع بيانات من مصادر الإحصاءات الاقتصادية والاجتماعية المتاحة مثل تعدادات السكان واحصاءات العمالة والأجور وإحصاءات التعليم واحصاءات الجرائم بالإضافة إلى نتائج البحوث والدراسات الخاصة التي تتم بأسلوب العينات مثل بحوث الإنفاق وميزانية الأسرة ، ثم يتم تحليل هذه البيانات واستخلاص النتائج المطلوبة . ويجب التأكد في جميع الحالات من سلامة أسلوب جمع البيانات وسلامة أساليب تحليلها .

مثال (٤) : دراسة الظواهر الطبيعية والتنبؤ بأنماطها في المستقبل . ويشمل ذلك ما يلي :

- كيف يمكن التنبؤ بأحوال الطقس ؟
- كيف يمكن تقدير عدد الأسماك من نوع معين في مياه الخليج ؟ وكيف

يمكن تقدير عدد الحيوانات البريه في الصحراء ؟

- كيف يمكن التنبؤ بالخصائص الوراثية للكائنات ؟
- كيف يتم تقدير الانتاج المتوقع من الخضروات والفواكه المختلفة خلال العام ؟

وتتطلب الإجابة عن مثل هذه الأسئلة بالطبع جمع البيانات المتاحة ثم استخدامها على نحو إحصائي سليم .

مثال (٥) : تطبيقات في مجالات الفنون والآداب : قد تستخدم الأساليب الإحصائية في مجالات الفنون والآداب . فمثلاً ، قد تستخدم هذه الأساليب لوصف أنماط الأعمال الفنية وصفاً رقمياً . وتستخدم هذه الأساليب ، تبعاً لذلك ، لتصنيف الأعمال الفنية وفي التعرف على مؤلفي القطع الفنية والأدبية مجهولة المؤلف .

مثال (٦) : تطبيقات قانونية : كثرت استخدامات الأسلوب الإحصائي لتقديم الإستشارات في قضايا قانونية في الآونة الأخيرة . فإذا تنازع طرفان بأن ادعى كل منهم حدوث واقعة معينة على نحو مختلف ، قد يقوم خبير إحصائي بحساب احتمال صدق كل طرف ويأخذ القاضي ذلك في الاعتبار عند إصدار حكمه .

لا عجب إذن ، مع هذا التعدد الواضح في مجالات تطبيق علم الإحصاء ، ملاحظة بعض الاختلافات بين المشتغلين في مجالات العلم المختلفة حول انطباعاتهم عن علم الإحصاء ، وذلك تبعاً لنوع الاستخدامات الأساسية للإحصاء في مجال كل منهم . فمثلاً :

- **في مجال الطب** : الاحصاء هو مجموعة الطرق التي تستخدم لاكتشاف العلاقات بين الأمراض ومسبباتها واكتشاف وتقويم قدرة أساليب العلاج المختلفة على شفاء الأمراض .
- **في مجال الاقتصاد وادارة الأعمال** : الإحصاء هو طرق عرض وتقديم الحقائق الرقمية واستخدامها كأساس لاتخاذ القرارات .

- **في مجال العلوم الزراعية** : الإحصاء هو طرق تصميم وتحليل التجارب الزراعية .
- **في مجال العلوم الهندسية والفيزياء والكيمياء** : الإحصاء هو طرق دراسة وتحليل الأخطاء التي تنشأ أثناء العمل نتيجة أخطاء في المقاييس أو عيوب فنية أخرى .
- **في مجال علوم الحياة** : الإحصاء هو طرق وصف وتفسير الاختلافات الطبيعية في الكائنات المختلفة .
- **في مجال علوم الاجتماع** : الإحصاء هو طرق استنتاج السمات العامة في البيانات الخاصة بالمجتمعات البشرية والتي تتميز بوجود اختلافات بين قيم المفردات المحتلفة .
- **في مجال الجغرافيا** : الإحصاء هو طرق وصف التوزيع الجغرافي للسكان ودراسة العلاقات بين المستوطنات البشرية والبيئة الطبيعية المحيطة بها .
- **في مجال التربية** : الإحصاء هو طرق وصف وتقويم القدرة على التعلم ، وقياس درجة الاستيعاب وبالتالي دراسة كفاءة البرامج التعليمية .
- **في مجال علم النفس** : الإحصاء هو طرق اكتشاف العلاقات السببية بناء على بيانات تجريبية يتم جمعها في مواقف مختلفة تتعلق بسلوك الكائنات الحية .

سنرى فيما بعد أن القاسم المشترك بين جميع هذه التطبيقات الاحصائية هو وجود مجموعة من الأساليب الرياضية التي تستخدم لأغراض التحليل . وبالتالي ، يحتاج المشتغلون بالإحصاء عادة إلى خلفية في الرياضيات . إلا أنه يجب الإشارة إلى أن معرفة الرياضيات غير كافية بمفردها لجعل الشخص قادراً على إجراء الدراسات الإحصائية في جميع مجالات التطبيق . ومن هنا يتضح أهمية قيام المشتغلين في مختلف أوجه العلم والمعرفة بتعلم المفاهيم الإحصائية الأساسية ، إذ يشجع ذلك على وجود رغبة متزايدة لديهم نحو استخدام الأساليب الإحصائية في دراساتهم وينمي قدراتهم على التعامل مع المتخصصين في الإحصاء لمساعدتهم في جمع وتحليل البيانات على أساس علمي سليم .

٣ ـ المفاهيم الأساسية في علم الاحصاء

يفيد العرض السريع للأمثلة العملية السابقة التي توضح مجالات التطبيق الاحصائي في تعريف المفاهيم الاحصائية الأساسية . وفيما يلي مناقشة موجزة لأهم هذه المفاهيم .

أ ـ جمع البيانات Data Collection

يدخل العمل الاحصائي بشكل أساسي في تصميم وتنفيذ عملية جمع البيانات وذلك بهدف ضمان الوصول إلى مستوى محدد للدقة والتأكد في نتائج هذه البيانات . ويشمل ذلك تعريف مفردات المجتمع محل الدراسة ، وتحديد نوع البيانات التي تجمع ، وعدد المفردات التي تجمع منها هذه البيانات ، وكيفية اختيار هذه المفردات ، ثم أسلوب الحصول على البيانات من كل منهم .

ب ـ الاحصاء الوصفي Descriptive Statistics

تتألف البيانات الاحصائية من قيمة (قراءة أو مشاهدة) واحدة أو أكثر لكل مفردة من المفردات التي تجمع منها البيانات . وتتميز البيانات الاحصائية بخاصية أساسية تتمثل في **إختلاف هذه القيم فيما بينها من مفردة إلى أخرى** . فمثلاً ، عند دراسة دخول الأسر في المدينة يلاحظ اختلاف هذه الدخول من أسرة لأخرى ، وعند دراسة أطوال نوع معين من الأشجار بعد عام من زراعتها يلاحظ وجود اختلافات في هذه الأطوال ، كذلك عند دراسة آراء السكان حول أمر معين يلاحظ اختلاف هذه الآراء من شخص لآخر وهكذا . وتجدر الإشارة إلى أن وجود هذه الاختلافات يعتبر أمراً عضوياً في جميع الظواهر المتعلقة بنشاط الانسان وخصائص بيئته . وتمثل هذه الاختلافات الأساس الذي يدعو لاستخدام الأساليب الاحصائية في الحياة العملية .

يبدأ التحليل الاحصائي لمجموعة من البيانات بمحاولة تقديم وصف لنمط الاختلاف السائد في هذه البيانات . ويشمل ذلك تنظيم وتلخيص وعرض

هـذه البيانـات وحساب عـدد من المقاييس الاحصـائية التي تعكس السمـات الأساسية لها . وتسمى مجموعة الطرق والأساليب التي تبحث في كيفية وصف نمط الاختلاف في مجموعة من البيانات بطرق الإحصاء الوصفي Descriptive Statistics .

جـ ـ المجتمع والعينة Population and Sample

يمكن أن يتم جمـع البيانـات الاحصائيـة بأحـد أسلوبين همـا أسلوب الحصر الشامل Complete Enumeration وأسلوب المعاينة Sampling . ويقصد بالحصر الشامل جمع بيانات من جميع المفردات المستهدفـة في الدراسـة ، على حين يقتصر الأمر في أسلوب المعاينة على جمـع بيانـات من بعض هذه المفـردات فقط . ويـلاحظ أنـه نتيجـة لـوجـود نمط إختـلاف في البيـانـات الاحصائية ، فإن القاعدة العامة لجمع البيانات تتمثل في جمع هذه البيانـات من أكبر عدد ممكن من المفردات بهدف استخدامها لاعطاء وصف جيد لنمط الاختـلاف . ويعتمد الاختيـار بين أسلوبي جمع البيـانات عـادة على ثـلاث اعتبارات متداخلة هي :

(١) **الموارد المتاحة لاجراء البحث**: ويقصـد بذلـك الموارد المـادية والفنية المتاحة بالاضافة إلى الفتـرة الزمنيـة المعطاة لاتمـام البحث . ويمكن القول كقاعدة عامة ، أن الموارد المتاحة لاجراء البحوث لا تسمح غالباً باجراء حصر شامل خاصة إذا كان عدد المفردات المستهدفة كبيراً أو إذا كانت البيانات المطلوبة على درجة عالية من التفصيل والتشعب . وتمثـل تعدادات السكـان التي تجريها الدول المختلفة بصفـة دورية استثنـاءاً لهذه القـاعدة ، ذلـك أن الدولة تضمن توفير الموارد اللازمة لاجراء التعداد لأسباب دستورية وقانونية .

(٢) **البيانات المتاحة**: في كثير من الحالات ، تكون البيانـات متاحـة فقط لجزء من المفردات المستهدفة في الدراسة . فمثـلاً ، إذا كان مـوضوع الدراسة يتعلق بـالأفراد المصـابين بمرض السكـر ، فإن مجمـوعة المفـردات المستهدفة في هـذه الحالـة هم المصابـون بهذا المـرض في كل مكـان وكل

زمان . ومع ذلك فإنه يمكن جمع بيانات فقط من أولئك المصابين حالياً والمقيمين في أماكن يمكن الوصول إليها . كذلك إذا كان البحث يهدف الى دراسة خصائص السلع التي ينتجها مصنع معين فإن مجموعة المفردات المستهدفة هي السلع التي ينتجها هذا المصنع الآن وفي المستقبل . في مثل هذه المواقف ، لا يمكن اجراء حصر شامل لجميع المفردات المستهدفة ولا بد من الاعتماد على بيانات غير كاملة (عينة) للوصول إلى نتائج تتعلق بمجموعة المفردات المستهدفة ككل .

(٣) نوع البيانات المطلوبة: قد تؤدي عملية الحصول على البيانات المطلوبة إلى القضاء على مفردات الدراسة . مثال ذلك فحص نوع معين من المصابيح الكهربائية لتقدير متوسط عمر المصباح حيث يتطلب قياس عمر المصباح استخدامه حتى يحترق ، أو فحص شحنة من البيض لتقدير نسبة الفاسد فيها . في مثل هذه الحالات ، لا بد من الاعتماد على بيانات غير كاملة تتعلق بجزء فقط من المفردات المستهدفة في الدراسة . تسمى المجموعة غير الكاملة للقيم أو المشاهدات التي يتم جمعها « بيانات العينة » . وتمثل هذه البيانات جزءاً من مجموعة أكبر من القيم تتعلق بجميع المفردات المستهدفة في الدراسة هي بيانات المجتمع .

تعريف (١) : المجتمع Population : يقصد بالمجتمع الاحصائي المجموعة الكاملة للقيم أو المشاهدات الخاصة بظاهرة ما والتي تتعلق بجميع المفردات محل الاهتمام في الدراسة . ويعني ذلك أن المجتمع هو هدف الدراسة الاحصائية وأن عملية البحث العلمي تؤدي في النهاية الى استخلاص نتائج تتعلق بهذا المجتمع .

تعريف (٢) : العينة Sample : هي مجموعة القيم أو المشاهدات التي تم جمعها فعلاً أثناء الدراسة ، والتي تمثل جزءاً من المجموعة الكاملة للقيم أو المشاهدات التي تمثل المجتمع .

يترتب على ما سبق الملاحظات الآتية :

١ - يختلف المعنى الاحصائي للمجتمع عن معناه اللغوي . ذلك أن المجتمع الاحصائي لا يعني مجموعة من الكائنات الحية ، وإنما هو مجموعة من الأرقام تمثل القيم أو المشاهدات التي تتعلق بظاهرة معينة لجميع المفردات المستهدفة في الدراسة . وقد لا تكون هذه المفردات أشخاصاً أوكائنات حية ، فمثلاً عند دراسة انتاج نوع جديد من القمح تكون المفردات هي نباتات القمح التي تزرع وعند دراسة خصائص السلع التي ينتجها مصنع معين تكون المفردات هي وحدات السلع المنتجة وهكذا .

٢ - يتم الاعتماد على نتائج العينة لدراسة خصائص المجتمع ولذلك فإنه من الضروري أن يتم اختيار العينة بأسلوب يسمح بذلك . ويتألف أسلوب اختيار العينة من عنصرين : الأول تحديد جميع العينات الممكن اختيارها من المجتمع والثاني وضع أسلوب يمكن من اختيار إحدى هذه العينات . يترتب على ذلك أن البيانات التي تجمع سوف تختلف من عينة لأخرى . وتسعى الأساليب الاحصائية لتصميم العينات إلى جعل هذه الاختلافات أقل ما يمكن . ومن ناحية أخرى ، نلاحظ أن هناك مجتمعاً واحداً بياناته ثابتة ومحددة .

٣ - بما أن المجتمع هو الهدف النهائي للدراسة ، فإنه يمكن النظر إلى العينة على أساس أنها تجربة تم اجراؤها للتعرف على خصائص المجتمع حيث تمثل بيانات العينة نتائج هذه التجربة . وتمثل عملية تعميم نتائج العينة إلى المجتمع الجزء الأكبر والأساسي من علم الاحصاء الحديث وتطبيقاته المتعددة . وتسمى الطرق التي تستخدم لهذا الغرض بطرق الاستنتاج الاحصائي Statistical Inference .

٤ - يتضح من ذلك أن هناك ثلاث مجالات للعمل الإحصائي هي :

(أ) طرق اختيار العينات وأساليب اجراء التجارب اللازمة لجمع البيانات .

(ب) طرق الاحصاء الوصفي .
(جـ) طرق الاستنتاج الاحصائي .

وتجدر الإشارة إلى أن هذه المجالات الثلاث متداخلة وتمثل نظاماً متكاملاً للتحليل الاحصائي . ويظهر ذلك من شكل (١) الذي يمثل مراحل العمل الاحصائي .

٥ ـ نخلص مما سبق إلى أن علم الاحصاء يكتسب أهميته في دراسة المشكلات في مجالات التطبيق المختلفة من أمرين :

الأول : أن الأساليب الاحصائية مصممة لدراسة الظواهر والمشكلات التي تختلف قيمها من مفردة لأخرى . وتجدر الإشارة أن هذا النوع من الظواهر والمشكلات يمثل الأغلبية الساحقة من الظواهر والمشكلات التي تنشأ في جميع أوجه الحياة العملية .

الثاني : أن الأساليب الاحصائية تسمح بالتعرف على شكل نمط الاختلاف في المجتمع محل الدراسة باستخدام عينة تمثل جزءاً فقط من هذا المجتمع .

د ـ المعاينة العشوائية **Random Sampling**

تعتمد أساليب الاستنتاج الاحصائي على نتائج العينة كأساس للتعرف على خصائص المجتمع . ويتطلب ذلك أن يتم اختيار مفردات العينة بحيث تكون العينة ممثلة بقدر الإمكان للمجتمع الذي سحبت منه .

ولما كان هدف اجراء البحث أساساً هو التعرف على خصائص المجتمع فإن اختيار عينة ممثلة تماماً للمجتمع أمر مستحيل عملياً نظراً لأن خصائص المجتمع غير معروفة قبل سحب العينة . لهذا السبب ، يلجأ الاحصائيون إلى محاولة الحصول على عينات « غير متحيزة » من خلال اتباع مبدأ العشوائية Randomness . ويقصد بالعشوائية إختيار مفردات العينة بأسلوب يسمح بإعطاء كل مفردة من مفردات المجتمع فرصة معلومة للظهور في العينة . ويتوقع في ظل هذا المبدأ أن تظهر الخصائص المختلفة للمفردات في العينة بنسب قريبة من نسب ظهورها في المجتمع .

شكل (١)

مراحل العمل الاحصائي

أساليب اختيار العينات
وتصميم التجارب

البيانات

تنظيم وتلخيص وعرض البيانات

الأشكال البيانية
الاحصاء الوصفي

تفسير وشرح النتائج

هل
البيانات من
عينة أم من
مجتمع

عينة

استنتاج احصائي

حصر شامل

نتائج وخلاصات تتعلق
بالمجتمع المستهدف من الدراسة

وتجدر الإشارة إلى أن أساليب الاستنتاج الاحصائي تصلح فقط لتعميم نتائج العينات العشوائية . أي أنه لا يمكن احصائياً استنتاج خصائص المجتمع بالاعتماد على عينات غير عشوائية . ويرجع ذلك إلى أن أسلوب العشوائية يسمح بإمكانية حساب الفروق المحتملة بين خصائص العينة وخصائص المجتمع وحساب درجة الثقة والتأكد في نتائج العينة وهو ما يمثل الأساس اللازم لتعميم نتائج العينة إلى المجتمع .

مثال : ترغب إدارة المكتبات في اختيار عينة من مائة شخص من المجتمع الطلابي في الجامعة لدراسة آراء الطلبة في اقتراح فتح المكتبة أيام الجمع . اقترح أحد الخبراء على الإدارة استخدام أحد الأسلوبين الآتيين :

الأول : الحصول على قائمة بأسماء جميع الطلبة والطالبات في الجامعة ، ثم اعطاء رقماً لكل إسم وكتابة كل رقم في ورقة صغيرة ثم وضع هذه الأوراق في كيس . ويتم اختيار مفردات العينة باختيار ١٠٠ رقم من هذا الكيس بعد خلط الأوراق جيداً قبل اجراء السحب .

الثاني : الذهاب إلى مطعم الجامعة وكتابة أسماء أول مائة طالب يحضرون لتناول طعام العشاء خلال يوم معين .

نلاحظ في هذا المثال أن الأسلوب الأول يضمن اختيار عينة عشوائية لأن كل مفردة من مفردات المجتمع لها نفس فرصة الظهور في العينة ، على حين أن الأسلوب الثاني يؤدي الى عينة غير عشوائية لأن كثيراً من مفردات المجتمع لا يكون لديها فرصة للظهور في العينة .

وتجدر الإشارة إلى أن في بعض الحالات ، ونتيجة لاعتبارات الصدفة ، قد تأتي العينة العشوائية غير ممثلة للمجتمع . إلا أن احتمال حدوث ذلك صغير ويمكن قياس حجمه ، كذلك فإن هذا الاحتمال يتناقص كلما كبر حجم العينة .

هـ ـ معلمة المجتمع واحصاء العينة Parameter and Statistic

تسمى أية خاصية من خصائص المجتمع معلمة Parameter ، وتسمى أية خاصية من خصائص العينة إحصاءاً Statistic . ولما كان هدف الدراسة الاحصائية دائماً هو التعرف على خصائص المجتمع فإن ذلك يعني دراسة معالم هذا المجتمع. وتعتمد أساليب الاستنتاج الاحصائي في هذا الصدد على الاحصاءات المشاهدة في بيانات العينة . وعادة ما يأخذ ذلك أحد شكلين ، تبعاً للأهداف المحددة للدراسة :

(١) تقدير معالم المجتمع : Estimation : مثال ذلك تقدير متوسط دخل الأسرة في الدولة في عام ١٩٨٥ أو تقدير نسبة الفاسد في انتاج أحد المصانع . . . الخ . ويتم ذلك بالاعتماد على الاحصاءات المتوافرة في العينة . فمثلاً عند تقدير نسبة الإناث في سكان أحد المجتمعات ، أخذت عينة عشوائية من ١٠٠٠ شخص وجد بينهم ٤٠٠ أنثى ، وبالتالي تكون نسبة الإناث في العينة = ٠,٤٠ ، وهذا هو الاحصاء المحسوب من بيانات العينة . يستخدم هذا الاحصاء لتقدير نسبة الاناث في المجتمع ككل (وهي المعلمة) . وللتعرف على مدى دقة هذا التقدير ، يمكن استخدام النظريات الاحصائية لتقدير حجم الخطأ المحتمل . فمثلاً قد يقال في هذا المثال أن الفرق بين الاحصاء المحسوب من العينة والنسبة الحقيقية في المجتمع لن يتعدى ٠,٠٢٧ « بدرجة ثقة = ٠,٩٥ » . وسنرى فيما بعد كيف يتم اجراء مثل هذه الحسابات .

(٢) اختبارات فروض تتعلق بمعالم المجتمع : Hypotheses Testing :
قد يكون الهدف من البحث هو اختبار ما إذا كان نوع جديد من المصابيح الكهربائية يعمر لمدة أطول من أعمار الأنواع المتاحة حالياً ، وقد يكون الهدف هو اختبار ما إذا كانت طريقة معينة لتدريس اللغة الانجليزية أفضل من الطرق الأخرى ، أو قد يكون الهدف هو اختبار ما إذا كانت هناك فروقاً بين الطلبة والطالبات في الجامعة في القدرة على التحصيل . . . الخ . في مثل هذه الحالات يتم تصميم تجربة (أو عينة) للحصول على البيانات اللازمة لاجراء الاختبار المطلوب ، ثم تستخدم أساليب الاستنتاج الاحصائي للاعتماد على الاحصاءات المحسوبة في العينة لاتخاذ القرارات المناسبة .

كمثال على ذلك ، تم اختبار فعالية مصل جديد للوقاية من شلل الأطفال باستخدام عينة كلية حجمها ٤٠ ألف تلميذ من تلاميذ المدارس الابتدائية . تم تقسيم هؤلاء التلاميذ الى مجموعتين متساويتين ، ثم طعم أفراد المجموعة الأولى بالمصل على حين ترك أفراد المجموعة الثانية بدون تطعيم . سجلت عدد حالات الإصابة بشلل الأطفال في المجموعتين فكانت ٣٣ في المجموعة

الأولى و ١١٠ في المجموعة الثانية . أي أن معدل الإصابة للأطفال الذين حصلوا على المصل كان تقريباً ربع المعدل للأطفال الذين لم يحصلوا عليه . وبناءاً على ذلك تقرر تعميم التطعيم بهذا المصل كجزء من حملة الوقاية ضد شلل الأطفال .

يلاحظ أن نتائج مثل هذا التحليل تأتي بناءاً على عينة ، وبالتالي قد تكون عرضة لأخطاء . إلا أن الأساليب الإحصائية توفر وسائل لقياس احتمالات حدوث هذه الأخطاء وتعمل على تقليل فرص حدوثها .

و ـ شرح وتفسير نمط الاختلاف في الظواهر الاحصائية :

Explaining Variability

لا يقتصر العمل الاحصائي على وصف نمط الاختلاف في الظواهر المختلفة ، وإنما يتعدى ذلك إلى محاولة تفسير وشرح أسباب هذه الاختلافات من خلال دراسة العوامل المختلفة التي قد تكون مرتبطة بها . ويتم ذلك بجمع بيانات عن العلاقات بين هذه العوامل وبين الظاهرة محل الدراسة ، ثم استخدام أساليب التحليل الاحصائي لدراسة هذه العلاقات . فمثلاً ، عند دراسة الاختلافات المشاهدة عند قياس أطوال مجموعة من الأطفال ، قد يمكن ارجاع هذه الاختلافات إلى الاختلافات في أعمار الأطفال أو إلى الاختلافات في أطوال آبائهم أو إلى الاختلاف في نظام تغذيتهم . . . الخ ، كذلك عند دراسة الاختلافات المشاهدة في أسعار السيارات المستعملة ، قد يرجع ذلك إلى الاختلافات في سنوات صنعها أو في طريقة استعمالها . . . الخ ، وهكذا .

وتمثل أساليب دراسة العلاقات بين الظواهر جزءاً أساسياً من الأسلوب العلمي لدراسة المشكلات ، ذلك أنه من الأهمية بمكان عند دراسة ظاهرة ما أن نتعرف على العوامل المختلفة التي تؤثر في هذه الظاهرة وتتأثر بها واستخدام هذه المعرفة كأساس لاتخاذ القرارات ورسم السياسات المختلفة .

تقدم النماذج الاحصائية الاطار اللازم للتوصل الى الطرق المختلفة للاستنتاج الإحصائي . ويقصد بالنموذج الاحصائي لظاهرة ما ، عملية تجريد رياضية يتم من خلالها تحديد الخصائص الأساسية للظاهرة ووصف ذلك رياضياً باستخدام نظرية الاحتمالات . وتستخدم هذه النماذج لدراسة الاحتمالات المناظرة للنتائج الممكنة في العينة وبالتالي تحديد درجة التأكد في نتائج العينة ، وهو ما يمثل الأساس الذي يتم بمقتضاه استخدام هذه النتائج لاستنتاج خصائص المجتمع .

فمثلاً ، عند أخذ عينة حجمها ٢٥ طالباً من بين طلبة الثانوية العامة والبالغ عددهم عشرة آلاف طالب بهدف دراسة توزيع الطلبة حسب التخصص (علمي / أدبي) ، يمكن تمثيل ذلك بصندوق به عشرة آلاف كرة متماثلة بعضها ذو لون أبيض (يمثل التخصص العلمي) وبعضها الآخر ذو لون أحمر (يمثل التخصص الأدبي) ، وتكون عملية اختيار العينة مماثلة لاختيار ٢٥ كرة من هذا الصندوق . مثل هذا النموذج البسيط يمكن استخدامه لحساب الاحتمالات المختلفة لتكوين مفردات العينة . أما إذا كان الأمر يتعلق بدراسة ظاهرة نوع المولود (ذكر/أنثى) فإنه يمكن استخدام نموذج يعتمد على نتائج رمي قطعة عملة حيث تشير الصورة الى الذكر وتشير الكتابة إلى الأنثى واستخدام ذلك لحساب الاحتمالات المختلفة لتوزيع الأطفال حسب النوع . كذلك عند اختبار ما إذا كان نوعاً معيناً من السماد يؤدي إلى زيادة ارتفاع طول الأشجار فإنه يمكن وضع نموذج يعتمد على العوامل التي تؤثر في عملية النمو ويصف النمط المتوقع لنمو الأشجار بدون استخدام السماد بحيث أنه إذا وجد أن أطوال الأشجار المسمدة لا تختلف عن هذا النمط المتوقع فإن ذلك يدل على عدم جدوى استخدام السماد والعكس .

ويلاحظ أن فائدة النموذج كإطار للحسابات الاحصائية تعتمد على قدرته على تمثيل الظاهرة الواقعية . إذ كلما كانت الخصائص الهامة للظاهرة ممثلة

في النموذج ، كلما قل احتمال الخطأ عند الاعتماد على نتائجه، والعكس .

ك ـ الاحصاء والحاسبات الآلية Statistics and Computers

يؤثر الانتشار الواسع للحاسبات الآلية تأثيراً بالغاً على أنماط الممارسة الاحصائية . ذلك أن هذه الحاسبات تساعد على الاسراع في تنفيذ الحسابات الاحصائية المعتادة من ناحية ، وتمكن الباحثين من إجراء حسابات ودراسات لم تكن ممكنة من قبل من ناحية أخرى . وقد تطورت الأساليب الاحصائية للاستفادة من هذا الوضع بحيث أصبحت استخدامات الحاسبات الآلية أمراً طبيعياً في مختلف التطبيقات الاحصائية. وتستخدم الحاسبات الآلية بشكل محدد في اجراء الآتي :

١ ـ تخزين كميات هائلة من البيانات الاحصائية بشكل يسهل معه استرجاعها واستخدامها وتحليلها كلما دعت الحاجة .

٢ ـ تخزين الأساليب الاحصائية اللازمة لتحليل هذه البيانات .

٣ ـ إجراء حسابات على درجة عالية من التعقيد بسهولة ويسر ، دون حاجة إلى تدخل انساني .

٤ ـ إجراء أنواع مختلفة من التحليل على نفس البيانات وبالتالي دراسة مدى كفاءة النماذج البديلة لوصف البيانات .

٥ ـ استخدام أساليب المحاكاة لتمثيل النماذج الاحصائية وبالتالي دراسة الخصائص المختلفة لهذه النماذج .

وتجدر الإشارة إلى أن هذا التطور قد أثر على أسلوب تدريس الاحصاء بحيث قل التركيز على طرق اجراء الحسابات المختلفة وزاد الاهتمام بكيفية التعامل مع البيانات واختيار استراتيجيات مناسبة للتحليل والقدرة على شرح وتفسير النتائج .

وتستخدم الحاسبات الآلية لاجراء جميع العمليات الاحصائية التي تناقش في هذا الكتاب . ويعتمد الباحثون في هذا الصدد على برامج للحساب

الآلي يعدونها بأنفسهم أو على مجموعات برامج متاحة سلفاً . ولعل أشهر هذه المجموعات وأكثرها استخداماً هي مجموعات SPSS ، SAS ، BMDP .

٤ ـ نبذة تاريخية

تشتق كلمة الاحصاء Statistics من الكلمة اللاتينية Status والتي تعني دولة .ذلك أن الأساليب الاحصائية قد استخدمت تاريخياً ولقرون عديدة في أغراض الحساب السياسي الذي يهتم بقياس قوة الدولة من خلال جمع وعرض البيانات الخاصة بالأوضاع السكانية والاقتصادية والسياسية لأفراد المجتمع . وقد استخدمت هذه البيانات كذلك لأغراض جمع الضرائب وفرض الخدمة العسكرية .

وتجدر الإشارة إلى أن فرعي التحليل الاحصائي (الاحصاء الوصفي والاستنتاج الاحصائي) قد اختلفا في نشأتهما التاريخية . فعلى حين ترجع جذور الاحصاء الوصفي الى ممارسات الحساب السياسي ، نجد أن الاستنتاج الاحصائي يرتبط أساساً بنظرية الاحتمالات التي نشأت وتطورت نتيجة الاهتمام بمسائل المقامرة ولعب الورق الذي كان سائداً خلال القرن السابع عشر .

ولعل أول دراسة احصائية معروفة هي تلك التي أجراها جرانت Graunt (١٦٢٠ ـ ١٦٧٤) والتي تتعلق بتحليل بيانات المواليد والوفيات في لندن بين عامي ١٦٠٤ ـ ١٦٦١ . وقد استفادت شركات التأمين من مثل هذه الدراسة فيما بعد لإنشاء جداول للحياة لفئات المجتمع المختلفة ، يتم على أساسها تحديد أقساط التأمين التي يدفعها الأفراد .

ومن ناحية أخرى ، بدأ التطور في نظرية الاحتمالات نتيجة اهتمام عدد من الرياضيين بمسائل المقامرة. ونخص بالذكر في هذا الصدد كل من باسكال Pascal (١٦٢٣ ـ ١٦٦٢) ودي فيرمات de Fermat (١٦٠١ ـ ١٦٦٥) اللذان لعبا دوراً أساسياً باكتشافهما قواعد مفيدة لحساب الاحتمالات استخدمت فيما

بعد كأساس لعمليات الاستنتاج الاحصائي . وقد قام كل من لابلاس Laplaee (١٧٤٩ ـ ١٨٢٦) وجاوس Gauss (١٧٧٧ ـ ١٨٥٥) بتطوير قواعد الاحتمالات وتطبيقها بشكل واسع في مجال الفلك . وتجدر الإشارة إلى أن جاوس هو الذي اقترح شكل التوزيع الطبيعي كنموذج للأخطاء التي تحدث عند اجراء التجارب . وقد استمر هذا التطور في نظرية الاحتمالات خلال القرنين التاسع عشر والعشرين وأدى إلى خلق نظرية متكاملة للاحتمالات .

وفي الوقت نفسه ، يعتبر كويتليت Quetelet (١٧٩٦ ـ ١٨٧٤) أول من طبق الأساليب الحديثة لجمع البيانات وقام بعقد المؤتمرات الاحصائية وأنشأ أول جمعية احصائية عالمية في بلجيكا أصبحت مثالاً يحتذى في البلدان الأخرى .

ويعتبر جالتون Galton (١٨٢٢ ـ ١٩١١) وپيرسون Pearson (١٨٥٧ ـ ١٩٣٦) أهم من ساهم في تطور الأساليب الاحصائية خلال القرن التاسع عشر . إذ قام جالتون باستخدام الأساليب الاحصائية لدراسة مسائل في علم الوراثة على حين قام پيرسون بإنشاء أساليب احصائية مبتكرة في مجالات متعددة . وقد ساهم كلاهما في ابتداع وتطوير أساليب دراسة العلاقات بين الظواهر الاحصائية المختلفة .

ولعل أكثر الاحصائيين شهرة في القرن العشرين هو فيشر Fisher (١٨٩٠ ـ ١٩٦٢) الذي قام بتطوير وابتداع كثير من الأساليب الاحصائية الحديثة . وتجدر الإشارة إلى أن كثيراً من هذه الأساليب قد نما من الحاجة لايجاد حلول لمشكلات عديدة نشأت في مجالات التجارب الزراعية والمقاييس في علم النفس وعلوم الحياة والاقتصاد والاجتماع والادارة والتربية وغيرها .

تمرينات

١ ـ حدد مشكلة أو ظاهرة في مجال تخصصك يمكن دراستها بالأسلوب الاحصائي .

٢ ـ ما هو المجتمع وما هي العينة في كل حالة من الحالات الآتية :

(أ) جمعت بيانات من ١٠٠ طالب من طلبة الجامعة عن رأيهم في كفاءة أحد أساتذة الاحصاء .

(ب) تمت دراسة انتاجية ٤٠ عاملاً من عمال إحدى الشركات لاختبار ما إذا كان العمال الذين يأخذون إجازاتهم بانتظام تزيد انتاجيتهم .

(حـ) تم قياس الانخفاض في ضغط الدم الناتج من استخدام دواء معين لعينة من ١٠٠ مريض من مرضى احدى العيادات .

٣ ـ ما هي نوع البيانات الاحصائية التي تعتقد بضرورة جمعها في الحالات التالية :

(أ) دراسة آراء طلبة الجامعة حول اقتراح بفتح المكتبات أيام الجمع .

(ب) دراسة ما إذا كان برنامج تدريب الموظفين قد أدى إلى رفع كفاءتهم .

(جـ) دراسة لأنماط الجرائم السائدة في الدولة .

٤ ـ لماذا يستخدم الاحصائيون مبدأ العشوائية عند اختيار العينات ؟

٥ ـ حدد في كل حالة من الحالات الآتية ما إذا كانت العينة عشوائية أم لا مع توضيح سبب اجابتك .

(أ) ترغب إحدى الصحف اليومية في التعرف على آراء السكان حول النظام المقترح للمواصلات العامة في المدينة . قامت الصحيفة بنشر استمارة بالبيانات المطلوبة يومياً لمدة أسبوع وطلبت من القراء ملأ هذه الاستمارات وإعادتها للصحيفة ، فجاء الرد من ١٥٠٠ قارئاً .

(ب) يريد مفتش الصحة التأكد من أن مصنع الألبان ينتج طبقاً للمواصفات الصحية المطلوبة . قرر المفتش زيارة المصنع في اليوم

الخامس من كل شهر لفحص عملياته .

(حـ) يراد اختيار عينة من ٢٠ أسرة من بين أسر الحي والبالـغ عددهـا ١٠٠ أسرة . كتبت أسماء الأسر على بطاقات متماثلة ثم خلطت هذه البطاقات جيداً واختير ٢٠ بطاقة من بينها .

٦ ـ حـدد مثـالا لحـالـة يؤدي الحصـول على البيـانـات فيهـا إلى تحـطيم المفردات .

٧ ـ أحيـاناً يعـرف علم الاحصاء بـأنه « علم اتخـاذ القرارات في ظـل عـدم التأكد » . اشرح المقصود بذلك .

٨ ـ ما هي المعلمة وما هو الاحصاء في كل حالة من الحالات التالية :

(أ) يرغب مدير الجامعة في كتابة تقرير يتعرض لمتوسط عدد الطلبة في كل مساق وكانت لديه المعلومات الآتية (i) في عينة من ٢٠ مسـاقاً وجد أن متوسط عدد الطلبة في كل مساق يساوي ٧٧ طـالباً ، (ii) في دراسة لجميع المساقات وجـد أن متوسط عـدد الطلبـة في كل مساق يساوي ٨٤ طالباً .

(ب) يراد قياس نسبة المدخنين في المجتمع ؛ أخذت عينة عشوائية من ٢٠٠ شخص في هذا المجتمع وجمعت بيـانات عن مـا إذا كانـوا مدخنين أم لا .

(حـ) يراد دراسة الزمن الذي يستغرقه الموظف في الوصول إلى عمله كل صباح . أخذت عينة عشوائية من ٥٠٠ موظف وسجل الزمن الخاص بكل منهم .

(ى) يراد التعرف على متوسط وزن حبة الطماطم المنتجة من نوع جديد من البذور . أخذت عينة من ٢٠٠ حبة من هذا النوع .

٩ ـ انـظر في صحيفتـك اليـوميـة ، وحـاول البحث عن تحقيق أو مـوضـوع يستخدم الاحصاء بشكل مفيد .

الباب الثاني

البيانات الاحصائية

١ ـ مقدمة

تسمى البيانات التي تجمع في بحث أو دراسة ما بإسم مجموعة البيانات الخاصة بتلك الدراسة Data Set . ويعتبر توافر مجموعة البيانات شرطاً أساسياً للدراسة الاحصائية لظاهرة ما ، ذلك أن التحليل الاحصائي للظاهرة لا يمكن أن يبدأ قبل جمع البيانات اللازمة وتنظيمها بأسلوب مفيد .

يعطي شكل (١) مثالاً لمجموعة بيانات عن بعض الخصائص لعينة من طلبة الجامعة ، حيث توجد بيانات عن النوع والجنسية والسنة الدراسية والمعدل العام وعدد أفراد الأسرة والمسافة المقطوعة يومياً في الذهاب الى الجامعة لكل طالب في العينة . ويفيد هذا المثال في إبداء الملاحظات الأولية التالية حول السمات العامة لمجموعات البيانات :

١ ـ تعطى مجموعة البيانات معلومات تتعلق بمفردات Elements مستهدفة في الدراسة ، حيث تجمع بيانات عن واحدة أو أكثر من هذه المفردات . ويمثل كل طالب في العينة في شكل (١) مفردة. وقد سبقت الإشارة إلى أن المفردات قد لا تكون بالضرورة أشخاصاً أو كائنات حية ، وذلك تبعاً لأهداف الدراسة .

٢ ـ تجمع البيانات عن خصائص المفردات المستهدفة في الدراسة . وتسمى كل خاصية من هذه الخصائص متغيراً إحصائياً Variable . ويلاحظ أن

المتغير الاحصائي هو ظاهرة تختلف قيمها من مفردة لأخرى وفق نمط معين ، مثال ذلك النوع والجنسية والسنة الدراسية والمعدل العام والمسافة المقطوعة يومياً في الذهاب الى الجامعة وعدد أفراد الأسرة في شكل (١) . ويكون الهدف من الدراسة الاحصائية هو التعرف على أنماط المتغيرات الاحصائية باستخدام البيانات المتاحة .

٣ ـ تتألف مجموعات البيانات من مشاهدات أو قراءات Observations عن المتغيرات الاحصائية محل الاهتمام للمفردات المستهدفة في الدراسة. وقد تكون مجموعة البيانات خاصة بمتغير واحد فقط Univariate Data Set، أو قد تكون شاملة لأكثر من واحد من هذه المتغيرات Multivariate Data Set .

٤ ـ قد تكون مجموعة البيانات خاصة بعينة أو بمجتمع وذلك تبعاً لعدد المفردات التي تجمع منها البيانات .

شكل (١)

مجموعة بيانات عن بعض خصائص طلبة الجامعة لعينة من عشرة طلاب

رقم الطالب	النوع	الجنسية	السنة الدراسية	المعدل العام	عدد أفراد الأسرة	المسافة بالكيلومترات
١	ذكر	مواطن	٤	١,٥٤	٥	٧,٥
٢	ذكر	غير مواطن	٤	٢,٥٧	٧	٢٩,٥
٣	ذكر	مواطن	٣	١,٧١	٨	٤
٤	انثى	مواطن	٢	٢,٤١	٦	١٥
٥	انثى	مواطن	٤	٣,٠٣	٤	٣,٥
٦	ذكر	غير مواطن	١	٢,٥٠	٦	٤
٧	انثى	مواطن	٢	٢,٩١	١٠	٤,٥
٨	ذكر	مواطن	٣	١,٥٠	٩	٢٥,٤
٩	انثى	غير مواطن	١	٢,٢٣	٧	٣,٥
١٠	ذكر	مواطن	١	٣,٠٩	٨	٢٩,٥

٢ ـ أنواع المتغيرات الاحصائية

سبقت الإشارة إلى أن هدف الدراسة الاحصائية هو التعرف على أنماط المتغيرات الاحصائية بناء على البيانات المتاحة . ويختلف أسلوب جمع وتحليل البيانات باختلاف أنواع المتغيرات التي تجمع عنها هذه البيانات . وفي هذا الصدد ، يتم التمييز بين أنواع المتغيرات المختلفة على أساس معنى وكيفية استخدام الأرقام للتعبير عن المشاهدات المأخوذة عن هذه المتغيرات . ويمكن تقسيم المتغيرات الاحصائية ، تبعاً لذلك ، إلى نوعين هما المتغيرات النوعية والمتغيرات الكمية .

أ ـ المتغيرات النوعية — Categorical or Qualitative Variables

وهي متغيرات يتألف كل منها من عدد من الأوجه المتنافية ، بحيث تقع كل مفردة من المفردات المستهدفة في الدراسة في وجه واحد من هذه الأوجه . مثال ذلك نوع الشخص (ذكر ، أنثى) ، جنسية الشخص (مواطن ، غير مواطن) ، محل الاقامة (ريف ، حضر) ، لون السيارة (أحمر ، أبيض ، أزرق ، . . .) ، التخصص الدراسي للطالب (رياضيات ، اقتصاد ، اجتماع ، هندسة ، . . .) ، الوضع الاجتماعي للأسرة (طبقة عليا ، طبقة متوسطة ، طبقة دنيا) ودرجة اصابة الفرد بالمرض (شديدة ، متوسطة ، ضعيفة) ، . . . الخ . وتكون عملية أخذ مشاهدات عن هذه المتغيرات مكافئة لتصنيف مفردات الدراسة على الأوجه المختلفة للمتغير . ويترتب على ذلك ، كما سنرى فيما بعد ، أن عملية تحليل هذه البيانات تعتبر أمراً سهلاً نسبياً وتعتمد أساساً على تحديد عدد مرات تكرار كل وجه من هذه الأوجه بين مفردات الدراسة .

قد يتميز المتغير النوعي بعدم وجود ترتيب منطقي بين الأوجه المختلفة التي يتألف منها . فمثلاً ، في حالة لون السيارة لا يوجد أساس لتمييز لون عن آخر ، وفي حالة دراسة التخصص الدراسي للطالب لا يوجد ترتيب لتفضيل تخصص عن آخر ، وفي حالة دراسة جنسية الشخص لا يمكن تمييز جنسية

عن أخرى ، وهكذا . ويسمى المتغير النوعي في هذه الحالة متغيراً تصنيفياً Nominal حيث تستخدم أوجه المتغير كأسماء أو أنواع لأغراض التصنيف فقط ، دون وجود علاقة ترتيبية بين هذه الأوجه .

وقد يوجد في حالات أخرى ترتيب منطقي بين أوجه المتغير النوعي . مثال ذلك الوضع الاجتماعي للأسرة (طبقة عليا ، طبقة وسطى ، طبقة دنيا) ودرجة موافقة الأفراد على الاقتراح الخاص بتقليل ساعات الدوام اليومي (موافق بشدة ، موافق ، معارض ، معارض بشدة) ، حيث يلاحظ أن الأوجه المختلفة للمتغير يمكن ترتيبها تصاعدياً أو تنازلياً . ويسمى المتغير النوعي في هذه الحالة متغيراً ترتيبياً Ordinal .

وتتضح الفروق بين البيانات التصنيفية والبيانات الترتيبية بملاحظة ما جرت عليه العادة من ترميز رقمي للأوجه المختلفة للمتغيرات النوعية وذلك لتسهيل التعامل مع البيانات على الحاسبات الآلية . فمثلاً ، عند دراسة متغير تصنيفي مثل نوع الشخص قد يعطي للوجه « ذكر » الرقم (صفر) وللوجه « أنثى » الرقم (١) ، حيث تستخدم هذه الأرقام كأسماء لأغراض التعريف فقط دون أن يكون لها معنى كمي . أما عند دراسة متغير ترتيبي مثل الوضع الاجتماعي للأسرة فقد تسمى الطبقة العليا «١» والطبقة المتوسطة «٢» والطبقة الدنيا «٣» . وفي هذه الحالة لا تستخدم الأرقام كمسميات فقط وإنما للدلالة أيضاً على الوضع النسبي لكل وجه بين الأوجه المختلفة للمتغير .

ب ـ المتغيرات الكمية **Numerical and Quantitative Variables**

في هذه الحالة ، يتم الحصول على المشاهدات من كل مفردة في شكل رقم أو كمية . وفي هذا الصدد ، يمكن أن نميز بين المواقف التي يتم الحصول فيها على هذا الرقم من خلال عملية عد Counting ، وتلك التي يتم فيها الحصول على الرقم من خلال أسلوب للقياس Measurement .

يسمى المتغير الكمي متغيراً متقطعاً Discrete إذا كان العد هو أسلوب الحصول على المشاهدات لهذا المتغير . مثال ذلك عدد أفراد الأسرة ، عدد

الغرف في المسكن ، عدد الحوادث اليومية عند تقاطع مروري معين ، عدد المساقات التي يدرسها الطالب ، عدد المصابيح التالفة في شحنة ما ، . . . الخ . وعلى ذلك فإن المتغير المتقطع هو المتغير الذي يأخذ قيماً من بين الأعداد (صفر ، ١ ، ٢ ، ٣ ، . . .) .

ويسمى المتغير الكمي متغيراً متصلاً Continuous إذا كان القياس هو أسلوب الحصول على المشاهدات ، مثل قياس عمر الشخص وقياس دخل الأسرة وقياس المسافة التي يقطعها الموظف في الذهاب إلى عمله كل صباح وقياس الزمن الذي يستغرقه الطالب للإجابة عن سؤال معين ، . . . الخ . ويلاحظ أن وحدات القياس في هذه الحالة يمكن أن تنقسم إلى عدد لانهائي من الأجزاء ، فالعمر يمكن أن يقاس لأي جزء من السنة ، والدخل لأي جزء من الدرهم ، والمسافة لأي جزء من الكيلومتر ، والزمن لأي جزء من الثانية ، وهكذا .

تختلف المتغيرات الكمية فيما بينها تبعاً لما إذا كانت المشاهدات الرقمية لها معنى كمي كامل أم لا . ويتحدد ذلك على أساس دلالة الرقم « صفر » كقيمة من قيم المتغير .

إذا كان الصفر نقطة تحكمية يمكن أن يختلف تعريفها من حين لآخر ، فإن البيانات تكون مبنية على مقياس البعد Interval- Seale Data . مثال ذلك درجات الحرارة ، والسنوات الهجرية والميلادية ، والمقاييس المختلفة لمستوى الذكاء ومستوى الصحة وقوة الزلازل ، . . . الخ .

أما إذا كان الصفر نقطة ثابتة ذات معنى مطلق ، فإن البيانات تكون مبنية على مقياس النسبة Ratio- Seale Data . مثال ذلك بيانات الطول والوزن والدخل والمسافة وكمية الوقود في السيارة . . . الخ .

إذا قيل أن « مع علياً صفراً من الدراهم » وأن « مع إبراهيم ضعف عدد الدراهم التي مع أحمد » فإن هذه العبارات لها معنى كمي كامل . أما إذا قيل أن درجة الحرارة تساوي صفراً مئوياً ، فهل يعني ذلك عدم وجود حرارة ؟

بالطبع لا ، إذ يمثل الصفر في هذه الحالة نقطة تحكمية متفق عليها وتناظر درجة تجمد المياه على مقياس درجة الحرارة . ومن الممكن بسهولة الاتفاق على نقطة أخرى لتمثل الصفر على المقياس . يترتب على ذلك أنه لا يمكن القول أن مستوى حرارة شيء درجة حرارته ١٠٠°م هو ضعف مستوى حرارة شيء آخر درجة حرارته ٥٠°م وإنما يمكن القول أن درجة الحرارة ٥٠°م تقع في منتصف المسافة بين درجة تجمد المياه (صفر°م) ودرجة غليان المياه (١٠٠°م) بمعنى أن كمية الطاقة اللازمة لرفع درجة حرارة شيء من صفر°م إلى ٥٠°م هي نفس كمية الطاقة اللازمة لرفع درجة حرارته من ٥٠°م إلى ١٠٠°م . يلاحظ أن المعنى الكمي للبيانات في هذه الحالة يسمح بتفسير البعد بينها ولا يسمح بتفسير النسبة . ويمكن ملاحظة نفس الشيء عند دراسة متغيرات أخرى مثل مستوى الذكاء للفرد أو مستوى الصحة في المجتمع ، وهي متغيرات لا يمكن قياسها مباشرة وإنما يتم استخدام اختبارات وجمع بيانات مساعدة تستخدم لإنشاء مقاييس للمتغير محل الدراسة . ولا يمثل الصفر في هذه المقاييس قيمة كمية ، وإنما هي نقطة تحكمية يتفق عليها .فإذا قيل أن مستوى الصحة في مجتمع ما يساوي الصفر فإن ذلك لا يعني انعدام الصحة في هذا المجتمع . ويتعين تبعاً لذلك توخي الحرص عند تحليل وتفسير مثل هذه البيانات وعند استخدامها لإجراء المقارنات المختلفة .

وتجدر الإشارة إلى أن أهداف الدراسة قد تتطلب معالجة متغير كمي كمتغير نوعي وذلك من خلال تجميع القيم المختلفة للمتغير الكمي والنظر إليها كأنواع . مثال ذلك تصنيف دخل الأسرة بين مستوى مرتفع ومستوى متوسط ومستوى منخفض . ويترتب على ذلك تقليل درجة التفصيل في البيانات مما يسهل عملية جمع البيانات وتحليلها .

٣ ـ المصادر الأولية والمصادر الثانوية للبيانات الإحصائية

سبقت الإشارة إلى أن التحديد الواضح لأهداف الدراسة الإحصائية لظاهرة ما هو الخطوة الأولى من خطوات هذه الدراسة . وتمثل تلك الأهداف الأساس اللازم للتعرف على نوع البيانات المطلوبة في الدراسة ، بالإضافة إلى

المصادر الممكنة لهذه البيانات وكيفية الحصول عليها وتجميعها .

ومن البديهي ، عند البدء في دراسة ما ، أن تقوم الهيئة المشرفة على البحث بالتعرف على الدزاسات وعمليات جمع البيانات التي تمت سابقاً في نفس المجال . ويشمل ذلك الدراسات والإحصاءات المنشورة بالإضافة إلى السجلات والدراسات الخاصة التي تحتفظ بها الهيئات والمؤسسات والشركات المختلفة . ويعتبر هذا العمل ضرورياً للتعرف على نوع المشكلات المتوقعة عند اجراء الدراسة وكيفية التغلب عليها ، وتحديد أية بيانات إحصائية متاحة يمكن الاستفادة منها ، هذا بالإضافة إلى اكتساب الخبرة اللازمة لإجراء الدراسة على أساس علمي سليم .

وتعتبر هذه الدراسات والبيانات النوع الأول من مصادر البيانات الإحصائية . وتوضح الأمثلة التالية بعض مجالات استخدام هذه المصادر :

(أ) إذا أريد المقارنة بين نفقة المعيشة في كل من القاهرة ولندن في عام ١٩٨٥ ، فإنه يمكن الاستعانة بالإحصاءات السنوية للأجور والأسعار في كل من المدينتين والتي تقوم الجهات الإحصائية المختصة بتجميعها ونشرها .

(ب) إذا كان هدف الدراسة هو التعرف على نمط التوزيع العمري والنوعي لسكان الدولة ، فإنه يمكن الاعتماد على بيانات تعداد السكان التي تقوم الجهات الإحصائية في الدولة بنشرها .

(حـ) إذا أريد دراسة أنماط الجريمة في بلد ما ، فإنه يمكن استخدام الإحصاءات التي تقوم جهات الأمن بتجميعها ونشرها .

(ء) إذا كان هدف الدراسة هو التنبؤ بأعداد الطلبة في المراحل التعليمية المختلفة بالدولة في عام ٢٠٠٠ فإنه يمكن الاستعانة بالإحصاءات والدراسات التي تعدها وزارة التربية بصفة دورية .

(هـ) إذا أريد التعرف على النمط الجغرافي لمبيعات سلعة ما ، فإنه يمكن استخدام سجلات الشركة التي تتولى توزيع هذه السلعة .

وينبغي عند استخدام المصادر الإحصائية المتاحة أن نفرق بين المصادر الإحصائية الأولية Primary Sourees وبين المصادر الإحصائية الثانوية Seeondary Sourees . وتحتوي المصادر الأولية على بيانات تكون جهة النشر هي مصدرها الأول (أي تكون هي الجهة المسؤولة عن جمع وإعداد هذه البيانات) . مثال ذلك مطبوعات تعداد السكان ، والكتاب الإحصائي السنوي للجامعة ، والإحصاءات السنوية لوزارة العمل ، والنشرة الإحصائية السنوية لوزارة التربية ، . . . الخ . أما المصادر الثانوية فتحتوي على بيانات تم تجميعها من مصادر مختلفة ، دون أن تكون جهة النشر مسؤولة عن إعدادها وجمعها . مثال ذلك الكتاب الإحصائي السنوي الذي تصدره الأجهزة المركزية للإحصاء في الدول المختلفة .

وتتميز كثير من المصادر الثانوية بتجميعها لأنواع مختلفة من الإحصاءات في مصدر واحد هذا بالإضافة إلى إمكانية استخدامها للتعرف على المصادر المختلفة لهذه البيانات . إلا أنه يفضل دائماً استخدام المصادر الأولية كلما كان ذلك ممكناً . ويرجع ذلك إلى أن البيانات في هذه المصادر تكون على درجة أعلى من التفصيل والشمول ، كما أنها تتميز بتوافر معلومات تتعلق بأساليب جمع البيانات والتعاريف والمفاهيم المستخدمة . هذا فضلاً عما هو معروف من أن نقل البيانات من مصدر لآخر قد يؤدي إلى حدوث أخطاء فيها سواء عن قصد أو غير قصد .

ويجب ، كقاعدة عامة ، توخي الحذر عند استخدام المصادر المختلفة للبيانات الإحصائية . إذ يتعين دراسة الظروف التي جمعت فيها البيانات ودراسة أنماط الاخطاء فيها والتعرف على مدى إمكانية استخدامها للوفاء بأغراض الدراسة . ويجب أن يشار دائماً إلى مصدر أي بيان إحصائي يستعان به عند إجراء دراسة ما .

. إذا كانت البيانات المطلوبة غير متوافرة في المصادر الإحصائية المتاحة ، فإنه لا بد من اللجوء إلى جمع هذه البيانات مباشرة من المفردات

المستهدفة في الدراسة . ونناقش في الفصل التالي العمليات الإحصائية المتعلقة بهذا الأسلوب .

٤ ـ جمع البيانات

من البديهي أن عملية جمع البيانات من مفردات الدراسة مباشرة تكون اكثر مشقة وأعلى تكلفة من الاكتفاء باستخدام المصادر الإحصائية المتاحة . إلا أن هذا الأسلوب يضمن للباحث الحصول على البيانات اللازمة لتحقيق أهداف الدراسة بشكل كامل . ونستعرض فيما يلي العناصر الأساسية لعملية جمع البيانات . وتجدر الإشارة إلى أن الإحاطة بهذه العناصر يفيد أيضاً في تقويم المصادر الإحصائية المتاحة واستخدامها تبعاً لذلك على النحو الأمثل .

أ ـ الدراسات التجريبية والدراسات غير التجريبية

Experimental and Nonexperimental Studies

هناك أسلوبين لجمع البيانات الإحصائية ، الأول هو الأسلوب التجريبي والثاني هو الأسلوب غير التجريبي أو أسلوب المسح أو المشاهدة . تعتمد الدراسة التجريبية على جمع البيانات باستخدام تجربة يتم تصميمها بأسلوب يسمح بالتحكم في العوامل المختلفة وقياس تأثير هذه العوامل على الظاهرة المستهدفة في الدراسة . ويستخدم الإحصائيون لهذا الغرض مبدأ العشوائية في تصميم التجارب بحيث تكون التجربة « غير متحيزة » وتمثل فيها جميع العوامل بشكل متوازن . فمثلا :

أ ـ عند إجراء المقارنة بين طريقتين لتدريس اللغة الانجليزية ، استخدمت عينة عشوائية من ١٠٠ طالب وزعت مفرداتها عشوائياً بالتساوي على الطريقتين . يلاحظ أن استخدام مبدأ العشوائية يهدف لتحقيق تكافؤ الفرص بين الطريقتين بأن تكون الخصائص المتوقعة للطلبة متشابهة في الطريقتين قبل البدء في إجراء التجربة .

ب ـ عند دراسة تأثير عقار معين على انقاص الوزن ، استخدمت عينة من ١٢

فأرا ، اختير ٦ من بينها عشوائياً لتعاطي هذا العقار ، بينما يستمر الباقون في نظامهم الغذائي العادي الذي لا يحتوي على هذا العقار .

حـ ـ عند إجراء المقارنة بين نوعين من السماد لدراسة تأثيرهما على نمو نوع ما من الأشجار ، استخدمت عشرة مناطق تحتوي كل منها على شجرتين . اختيرت إحدى الشجرتين في كل منطقة عشوائياً للمعالجة بالنوع الأول من السماد بينما عولجت الشجرة الأخرى بالنوع الثاني .

ومن ناحية أخرى ، يقتصر الأمر في الدراسة غير التجريبية على ملاحظة أو مشاهدة أو مسح الوضع القائم فعلاً في ظاهرة ما ، دون أن يكون هناك تحكم في العوامل المختلفة أو تدخل في الظروف التي تؤدي إلى هذه البيانات . فمثلاً :

أ ـ عند دراسة العلاقة بين عدد أفراد الأسرة ونسبة ما تدخره الأسرة من دخلها في مجتمع ما ، يتم جمع بيانات من كل أسرة عن عدد أفرادها وعن نسبة ما تدخره من الدخل . ويلاحظ عدم إمكانية استخدام الأسلوب التجريبي في هذه الحالة ، إذ لا يعقل إجبار الأسر على تحديد عدد أفرادها مسبقاً ثم دراسة تأثير ذلك على الإدخار .

ب ـ عند دراسة تأثير نظام معين للتغذية على إطالة عمر الأفراد ، لا يمكن إجبار أي فرد على اتباع هذا النظام طيلة حياته وإنما لا بد من البحث عن مجتمعات يتبع أفرادها هذا النظام بشكل طبيعي ثم ملاحظتهم .

حـ ـ عند تقويم كفاءة أحد البرامج التدريبية للعاملين في إحدى المؤسسات ، دعي العاملون للإلتحاق بالبرنامج على أساس تطوعي . ثم تمت مقارنة كفاءة هؤلاء العاملين بعد انتهائهم من البرنامج مع كفاءة العاملين الآخرين . يلاحظ أن هذا الأسلوب في إجراء الدراسة هو اسلوب غير تجريبي بسبب غياب مبدأ العشوائية عند اختيار المتدربين .

ء ـ لا يمكن بشكل عام اتباع الأسلوب التجريبي عند دراسة المتغيرات

المختلفة التي تنشأ في مجالات العلوم الاجتماعية مثل الدخل والانفاق وأنماط الجريمة ورأي الشخص في أمر معين . . . الخ . ولا بد في هذه الحالات من الاعتماد على البيانات التي تجمع من خلال مسح الوضع القائم في هذه المتغيرات .

ويلاحظ أن كلاً من الدراسات التجريبية والدراسات غير التجريبية يستخدم بشكل فعال في دراسة الظواهر الإحصائية المختلفة . وإن كانت الدراسات التجريبية تؤدي إلى الحصول على بيانات اكثر قوة لدراسة العلاقات المختلفة بين الظواهر ، وبصفة خاصة يعتبر الأسلوب التجريبي أكثر ملاءمة لدراسة علاقات السببية بين المتغيرات المختلفة . ولذا ينصح دائماً باستخدام هذا الأسلوب في المواقف التي تسمح بذلك .

ب ـ طرق الحصول على البيانات من المفردات

Data - Acquisition Procedures

تتعدد طرق الحصول على البيانات من مفردات الدراسة ، سواء استخدم الأسلوب التجريبي أم أسلوب المسح لجمع البيانات . وتعتبر طريقة المقابلة الشخصية وطريقة الاستبيان الذاتي وطريقة الملاحظة اكثر هذه الطرق شيوعاً في الاستخدام .

تعتمد طريقة المقابلة الشخصية Personal Interview على قيام جامع البيانات بتوجيه عدد من الأسئلة تكون متضمنة في صحيفة للبحث ، إلى كل مفردة من مفردات الدراسة ثم تسجيل إجابات الأفراد عنها في الأماكن المخصصة لها . فمثلاً :

أ ـ عند إجراء تعداد السكان ، يقوم العداد بمقابلة رب كل أسرة والحصول منه على البيانات المطلوبة عن أسرته .

ب ـ في دراسة عن خصائص سوق العمل في الدولة ، يقوم جامعو البيانات بمقابلة أصحاب الأعمال والحصول على البيانات المطلوبة من كل منهم .

حـ ـ في بحوث التسويق ، يقوم جامعو البيانات بمقابلة المستهلكين والحصول منهم على بيانات تتعلق بآرائهم حول خصائص السلع المختلفة .

ء ـ في دراسة عن أسباب تخلف الطلبة عن حضور المحاضرات ، قام جامعو البيانات بالحصول على البيانات من خلال المقابلة المباشرة للطلبة المستهدفين في الدراسة .

وتضمن المقابلة المباشرة بين جامعي البيانات ومفردات الدراسة قيام جامع البيانات بقراءة وتوضيح معنى الأسئلة لمفردات الدراسة . ويؤدي اتباع هذه الطريقة إلى ارتفاع في معدلات استجابة هذه المفردات . ويكتسب هذا الأمر أهمية خاصة في المواقف التي يرتفع فيها مستوى الأمية أو يقل مستوى الوعي الإحصائي بين مفردات المجتمع . ومن ناحية أخرى ، قد يترتب على المقابلة الشخصية وجود تحيز في البيانات ، ناشىء عن عدم خبرة أو عدم أمانة جامعي البيانات . اذ قد يهمل جامعو البيانات اتباع التعليمات المعطاة لهم أو قد يتعاملون مع مفردات الدراسة بأسلوب يؤثر في نوعية إجاباتهم أو قد يخطئون عند تسجيل البيانات المعطاة لهم .

أما طريقة **الاستبيان الذاتي** Self- Enumeration فيتم فيها توفير صحيفة البحث لكل مفردة من مفردات الدراسة مع وضع تعليمات واضحة عن كيفية قيامهم بالإجابة عن الأسئلة المتضمنة فيها. فمثلاً :

أ ـ في دراسة عن خريجي الجامعة ، تم ارسال صحيفة البحث لكل خريج وذلك بهدف الحصول على إجابات عن أنشطتهم المختلفة منذ التخرج .

ب ـ عند تسجيل السيارات في إدارة المرور ، يقوم مالك كل سيارة بملأ صحيفة تتضمن بيانات عن نوع السيارة وتاريخ صنعها وخصائصها المختلفة .

حـ ـ قامت الجمعية الإحصائية بإرسال صحيفة استبيان لكل عضو من أعضاء

الجمعية بهدف استطلاع آرائهم حول محتويات المجلة الشهرية التي تصدرها الجمعية .

د - تقوم شركات الأدوية بإرسال صحف استبيان للأطباء بين الحين والآخر وذلك للحصول على بيانات تتعلق بآراء وملاحظات الأطباء حول الأدوية التي تنتجها هذه الشركات .

ويلاحظ أن استخدام هذه الطريقة يؤدي إلى جمع البيانات بتكلفة أقل ، كما أنه يتفادى أخطاء التحيز التي قد تنشأ من تدخل جامعي البيانات . ولا يمكن استخدام هذه الطريقة إذا كان مستوى الأمية مرتفعاً بين مفردات المجتمع . وتتسم طريقة الاستبيان الذاتي عموماً بانخفاض معدلات الاستجابة بين المفردات ، بما يترتب على ذلك من تحيز في البيانات التي تجمع . وينشأ هذا التحيز بسبب كون المستجيبين عينة غير ممثلة للمفردات المستهدفة في الدراسة . ويفسر ذلك حرص هيئات جمع البيانات على رفع معدلات استجابة المفردات عن طريقة المتابعة والاتصال المتتالي بمفردات الدراسة وحثها على ملأ الاستبيان وإعادته .

وتعتمد طريقة الملاحظة **Observation** على المشاهدة أو الفحص المباشر لظاهرة ما ثم تسجيل نتيجة هذا الفحص . فمثلاً :

أ - عند دراسة نسبة الإصابة بمرض العيون بين تلاميذ المدارس الابتدائية ، يقوم الطبيب بفحص كل تلميذ وتسجيل نتيجة هذا الفحص .

ب - في دراسة عند حركة المرور عند تقاطع ما ، يتم ملاحظة عدد السيارات المستخدمة لهذا التقاطع خلال فترات زمنية محددة .

حـ - يتم في الاختبارات النفسية المختلفة ملاحظة وتسجيل رد فعل مفردات الدراسة لمؤثرات معينة .

د - في عمليات ضبط الانتاج بالمصانع المختلفة ، يتم قراءة درجات حرارة الأجهزة المختلفة بين الحين والآخر وتسجيلها ، وذلك للإحاطة بأنماط التغير في حرارة هذه الأجهزة .

وتتميز هذه الطريقة ببساطتها وبأنها تؤدي إلى الحصول على البيانات المطلوبة بشكل مباشر ، مع التقليل من الأخطاء التي قد تحدث عند جمع البيانات . وتعتمد كفاءتها على مدى « عدم تحيز » الأشخاص الذين يقومون بالملاحظة ، وبالتالي يجب أن يكون هؤلاء الأشخاص على مستوى مقبول من الخبرة ومدربين تدريباً جيداً .

حـ ـ إعداد صحيفة البحث Questionaire Design

يعتمد نجاح عملية جمع البيانات بشكل أساسي على جودة صحيفة البحث المستخدمة لجمع هذه البيانات . ونناقش فيما يلي بإيجاز بعض الاعتبارات الهامة التي يجب مراعاتها عند إعداد صحيفة البحث . وتجدر الإشارة إلى أن نوعية الأسئلة المستخدمة في صحيفة البحث قد تختلف باختلاف طريقة جمع البيانات ، إذ يمكن في حالة استخدام أسلوب المقابلة الشخصية إدراج أسئلة اكثر تشعباً واكثر تعقيداً .

١ ـ **ضرورة ترتيب الأسئلة بعناية**: يجب أن تحتوي صحيفة البحث على الأسئلة ذات العلاقة المباشرة بأهداف الدراسة فقط . وينبغي دائماً البدء بالأسئلة الخاصة بتعريف المفردة ، ثم تأتي بعد ذلك بعض الأسئلة البسيطة عن المتغيرات محل الدراسة ، وتؤجل الأسئلة الصعبة أو المثيرة للجدل حتى النهاية . يجب أيضاً مراعاة التسلسل المنطقي في ترتيب الأسئلة ، وإذا كان البحث يتعلق بعدة أمور فينبغي الانتهاء من الأسئلة الخاصة بأمر ما قبل البدء في الأسئلة الخاصة بأمر آخر . ويمكن استخدام أكثر من سؤال عن نفس الموضوع للتأكد من صحة الإجابات ، إذا دعت الحاجة إلى ذلك .

٢ ـ **صياغة الأسئلة بإيجاز وبلغة واضحة**: يجب أن لا يكون عدد الأسئلة كبيراً ، كما ينبغي أن يتعلق كل سؤال بفكرة واحدة فقط . ويجب أن يوضع السؤال بشكل واضح بحيث لا يختلف مفهومه من شخص لآخر . وفي هذا الصدد يفضل أن تحدد الإجابات المختلفة الممكنة عن كل سؤال ، بحيث يختار الفرد إحدى هذه الإجابات . فمثلاً :

أ ـ إذا سئل الفرد عن حالته الزواجية ، فقد يطلب منه الاختيار بين :

☐ لم يسبق له الزواج ☐ متزوج ☐ أرمل ☐ مطلق .

ب ـ إذا سئل الفرد عما إذا كان ينوي قضاء الصيف بالخارج ، فإن الإجابات الممكنة هي : ☐ نعم ☐ لا

حـ ـ إذا سئل الطلبة المسجلين في معاهد التعليم المختلفة عن عدد السنوات الدراسية اللازمة للحصول على شهادة اتمام الدراسة ، فقد يطلب منهم الاختيار بين :

☐ عامين أو أقل ☐ أكثر من عامين وأقل من ٤ أعوام

☐ ٤ أعـوام أو أكثر .

ويجب كقاعدة عامة ، الابتعاد عن الأسئلة المفتوحة التي يترك للفرد فيها الإجـابة عنهـا بلغته الخـاصة ، وذلـك لصعـوبـة تفسيـر وتصنيف مثـل هـذه الاجـابات . وتجـدر الإشارة إلى أن هـذا النوع من الأسئلة قـد يستخـدم في المراحل الاستطلاعية للتعرف على الأنماط المتوقعة للإجابات ، ثم استخدام ذلك في صياغة الأسئلة في شكلها النهائي .

ويجب كذلك تفادي استخدام الأسئلة التاريخية وتلك التي تعتمد على الذاكرة ، نتيجة ما قد يؤدي ذلك إلى تحيز وأخطاء في البيانات .

٣ ـ التعريف الدقيق للمفاهيم المستخدمة : يجب شرح معنى التعاريف والمفاهيم ووحدات القياس المستخدمة في صحيفة البحث شرحاً دقيقاً بحيث لا يكـون هنـاك غمـوضـاً أو لبس . فمثـلاً عنـد السؤال عن عـدد حجـرات المسكن ، يجب شرح المقصود بالحجرة ، وعنـد السؤال عن الأجر ، يجب تحديد ما إذا كان ذلك هو الأجر اليومي أو الأسبوعي أو الشهري وهكذا .

٤ ـ تلافي الأسئلة الحساسة والمثيرة للغضب : مثـال ذلـك الأسئلة المتعلقة بالسلوك الشخصي للأفراد . ذلـك أن استخدام هـذه الأسئلة بشكل مباشر قد يؤدي إلى الحصول على إجـابات مضللة أو قـد يؤدي إلى إغضاب الأفراد وبالتالي إلى انخفاض معدلات استجابتهم للدراسة .

٥ ـ الابتعاد عن العوامل التي قد تسبب التحيز في الإجابات : وفي هذا الصدد ، يجب أن تصاغ الأسئلة بأسلوب حيادي لا يؤدي إلى الإيحاء بإجابات معينة لمفردات الدراسة . كذلك يجب التأكد من أن الأسئلة تسعى للحصول على إجابات تعتمد على حقائق واقعة وليس على انطباعات الأفراد التي قد تكون غير دقيقة أو غير قابلة للمقارنة .

٦ ـ إعداد تعليمات واضحة للإجابة عن الأسئلة :يجب أن تكون هناك تعليمات واضحة لجامعي البيانات عن كيفية إجراء المقابلات الشخصية والحصول على البيانات المطلوبة . ويجب أن يختار جامعي البيانات بعناية وأن يتم تدريبهم وتوعيتهم بأهمية البحث وبأساليب اكتساب تعاون مفردات الدراسة .

ويكتسب وجود هذه التعليمات أهمية خاصة عند استخدام أسلوب الاستبيان الذاتي لجمع البيانات . ويجب في هذه الحالة إرفاق هذه التعليمات مع صحيفة البحث بالإضافة إلى خطاب يشرح أهداف البحث وأهميته للمصلحة العامة ، وذلك بهدف رفع معدلات استجابة المفردات للدراسة .

د ـ أساليب المعاينة Sampling Techniques

سبقت الإشارة إلى أن جمع البيانات الإحصائية يمكن أن يتم باستخدام أسلوب الحصر الشامل أو أسلوب المعاينة . ويفضل استخدام عينة في معظم المواقف لأسباب متعددة منها عدم ملاءمة المجتمع لإجراء حصر شامل ، وعدم كفاية الموارد المادية والفنية المتاحة لإجراء البحث ، وضيق الفترة الزمنية المعطاة لإتمام الدراسة ، وإمكانية استخدام العينات للحصول على بيانات على درجة عالية من التفصيل . هذا فضلاً عن أن أساليب الاستنتاج الإحصائي تمكن من تعميم نتائج العينة إلى المجتمع الذي سحبت منه .

هناك ثلاث أمور أساسية يتعين تحديدها إذا ما تقرر استخدام أسلوب المعاينة لجمع البيانات الاحصائية ، وهي :

١ ـ تحديد المجتمع المستهدف في الدراسة Target Population ، ويقصد بذلك مجموعة المفردات التي يراد دراستها واستنتاج خصائصها بناءاً على نتائج العينة . ويطلق على القائمة التي تحتوي على جميع مفردات هذا المجتمع إسم اطار المعاينة Sampling Frame . ويعتبر إعداد الإطار أمراً ضرورياً لاختيار العينة . ويجب أن يكون الإطار جيداً بقدر الإمكان بحيث يشمل جميع مفردات المجتمع دون تكرار . ويمثل ذلك إحدى المشكلات الهامة في أسلوب المعاينة نتيجة نقص المعلومات من ناحية والتغير المستمر في تركيبة المجتمعات من ناحية أخرى . ولا تكون العينة ممثلة للمجتمع اذا كانت هناك اختلافات بين الإطار والمجتمع المستهدف . مثال ذلك استخدام دليل الهاتف كإطار لاختيار عينة من سكان إحدى المدن ، أو استخدام نتائج تعداد السكان الذي أجري منذ خمس سنوات كإطار لاختيار عينة من الأسر في الدولة .

٢ ـ تحديد المتغيرات التي تجمع عنها البيانات في العينة . مثال ذلك رأي الشخص في أمر معين ، المستوى التعليمي للشخص ، دخله الشهري ، . . . الخ . وتتحدد هذه المتغيرات في ضوء التحديد الدقيق لأهداف الدراسة .

٣ ـ تحديد نوع العينة المستخدمة في الدراسة . وفي هذا الصدد ، يجب التمييز بين نوعين من العينات : الأول هو العينات العشوائية Random Samples ، والثاني هو العينات التحكمية Judgement Samples . ويقصد بالعينات العشوائية تلك التي يتم اختيارها بأسلوب يسمح بتحديد احتمال ظهور كل مفردة من مفردات المجتمع في العينة ، وهي عينات يمكن استخدامها لأغراض الاستنتاج الإحصائي . ويسمى هذا النوع من العينات أحياناً بالعينات الاحتمالية Probability Samples أو العينات الإحصائية Statistieal Samples .

أما العينات التحكمية فهي عينات غير احتمالية يتم اختيارها وفق معايير يحددها الباحث ويعتقد أنها تؤدي إلى الحصول على عينة « ممثلة »

للمجتمع . وينصح كقاعدة عامة بتفادي هذا النوع من العينات كلما كان ذلك ممكناً لأنه لا يمكن استخدام الأساليب الاحصائية لتعميم نتائجها إلى المجتمع ، هذا بالإضافة إلى ما قد تعانيه هذه العينات من تحيز ناتج عن عدم كفاءة المعايير التي تستخدم كأساس لاختيارها .

وتجدر الإشارة إلى أن كثيراً من العينات قد تصمم أساساً كعينات احتمالية ، ولكنها تصبح عينات غير احتمالية بسبب المشكلات التي تنشأ عند جمع البيانات . مثال ذلك ما يحدث عند استخدام أسلوب الاستبيان الذاتي لجمع البيانات من مفردات عينة احتمالية . إذا اهتم جزء فقط من مفردات العينة بملأ الاستبيان وإعادته ، وتقرر اعتبار هؤلاء كعينة ممثلة للمجتمع المستهدف ، فإن هذه العينة تكون عينة غير احتمالية .

ونشير فيما يلي بإيجاز إلى بعض أنواع العينات العشوائية الهامة .

١ ـ العينة العشوائية البسيطة Simple Random Sample

وهي عينة يتم اختيارها بطريقة تسمح بإعطاء جميع العينات الممكنة من المجتمع ، والتي لها نفس الحجم نفس الفرصة في الاختيار . فمثلاً إذا كان لدينا مجتمعاً مكوناً من ٦ أفراد هم أ ، ب ، حـ ، د ، هـ ، و ، ويراد اختيار عينة حجمها ٢ من هذا المجتمع ، يلاحظ أن جميع العينات الممكنة والتي حجم كل منها = ٢ في هذه الحالة هي :

أب ، أحـ ، أد ، أهـ ، أو ، ب حـ ، ب د ، ب هـ ، ب و ، حـ د ، حـ هـ ، د هـ ، د و ، هـ و .

إذا تم اختيار إحدى هذه العينات بحيث يعطي لكل منها نفس الفرصة في الاختيار ، فإن العينة الناتجة تكون عينة عشوائية بسيطة .

إذا كان حجم المجتمع كبيراً نسبياً فإنه يكون من المستحيل رصد جميع العينات الممكنة ثم إختيار إحداها كما في المثال السابق . ويعتمد الاحصائيون في هذه الحالات على جداول تسمى جداول الأرقام العشوائية لاختيار العينة من الإطار مباشرة .

ويلاحظ أن إعطاء نفس الفرصة لجميع العينات في أسلوب المعاينة العشوائية البسيطة يرجع سببه إلى أن المعلومات المتاحة في هذه الحالة تحتم افتراض أن جميع مفردات المجتمع متجانسة وبالتالي تكون كل عينة من العينات الممكنة ممثلة للمجتمع ولا يوجد ما يدعو لتفضيل إحداها عن الأخرى .

٢ ـ العينة العشوائية الطبقية **Stratified Random Sample**

إذا توافرت معلومات إضافية عن مفردات المجتمع تدل على عدم تجانس هذه المفردات بالنسبة للمتغيرات محل الدراسة ، فإنه يمكن استخدام هذه المعلومات لتصنيف مفردات المجتمع إلى مجموعات أو طبقات متجانسة ثم تختار عينات عشوائية بسيطة مستقلة من هذه الطبقات .

فمثلاً ، إذا كان معلوماً في المثال السابق أن الأفراد أ ، حـ ، د ذكوراً وأن ب ، هـ ، و إناثاً وأريد اختيار عينة حجمها ٢ من هذا المجتمع بحيث يمثل فيها كل من الذكور والإناث فإنه يلاحظ أن العينات الممكنة في هذه الحالة هي :

أب ، أهـ ، أو ، بحـ ، حـهـ ، حـو ، دب ، دهـ ، دو وتختار إحداها كعينة عشوائية طبقية .

ويلاحظ أن العينة العشوائية الطبقية تعتمد على المعلومات المتوافرة عن مفردات المجتمع بشكل أفضل من اعتماد المعاينة العشوائية البسيطة وبالتالي يتوقع أن تكون العينة العشوائية الطبقية أكثر دقة من العينة العشوائية البسيطة وذلك لقرب خصائصها من خصائص المجتمع . هذا فضلاً عن أن المعاينة الطبقية توفر بيانات يمكن استخدامها لدراسة خصائص كل طبقة من طبقات المجتمع على حدة .

٣ ـ العينة العشوائية المنتظمة **Systematic Random Sample**

تختار هذه العينة بحيث تكون مفرداتها موزعة بشكل منتظم على إطار المعاينة . ويتم اختيار العينة بتقسيم مفردات المجتمع إلى فئات متساوية

عـددهـا يسـاوي حجم العينـة ، ثم تختـار إحـدى مفـردات الفئـة الأولى عشوائياً . وتتألف العينة من هذه المفردة والمفردات المناظـرة لها في الفئـات الباقية . فمثلاً عند إختيار عينة عشوائية منتـظمة حجمهـا ٢ من المجتمع أ ، ب، حـ ، د، هـ ، و ، نبـدأ بتقسيم هذا المجتمـع إلى فئتين همـا (أ ، ب ، حـ) ، (ء ، هـ ، و) . ثم تختار إحدى مفردات الفئة الأولى عشـوائياً فـإذا ظهرت أ كانت العينة هي (أ ، د) ، وإذا ظهـرت ب كانت العينة هي (ب ، هـ) ، وإذا ظهرت حـ كانت العينة هي (حـ ، و) أي أن العينات العشـوائية المنتظمة في هذه الحالة هي (أ د ، ب هـ ، حـ و) .ويلاحظ أن العينة الناتجة عشوائية لأن المفردة الأولى تختار عشوائياً ، ومنتـظمة لأن المفـردات الأخرى تختار بانتظام تبعاً لذلك .

ويمتاز هذا النوع من العينات بسهولة التنفيذ خاصة اذا كان اطار المعاينة يأخذ شكل السجل المنظم . ويمكن النظر للعينة العشوائية المنتظمة كتقريب للعينة العشوائيـة البسيطة إذا كـانت مفردات الإطـار غير مـرتبة داخـل الفئات بشكل يرتبط مع البيانات المراد جمعها . ويؤدي وجود مثل هذا التـرتيب إلى انخفاض كفاءة العينة لأن مفرداتها في هذه الحالة ستكون متشابهة مما يعني أن العينة غير ممثلة بخصائص المجتمع وتكون كمية المعلومات فيها محدودة .

٤ ـ العينة العنقودية Cluster Sample

يقسم المجتمع في هذه الحالة إلى عدد من المجموعات يسمى كل منها عنقوداً ، بحيث يحتوي كل عنقود على عدد من مفردات المجتمع . ثم تختار عينة من هذه العناقيد وتجمع البيانات من جميع مفرداتها . فمثـلاً ، إذا أريد اختيار عينة عشوائية عنقودية من طلبة المدارس الابتدائية في الدولة فإنه يمكن النظر إلى الفصول الدراسية المختلفة كعناقيد ، وتختار عينة من هذه الفصول الدراسية ثم تجمع البيانات المطلوبة من الطلبة في هذه الفصول .

ويتميز هذا الأسلوب عن أسلوب المعاينة العشـوائية البسيـطة بأنـه أقل تكلفة وأسهل في التنفيذ من ناحية ، وأنه يتلافى مشكلة إعداد إطـار المعاينـة

لجميع مفردات المجتمع من ناحية أخرى . إذ أن تطبيق هذا الأسلوب يحتاج إلى إطار للعناقيد فقط وليس إلى إطار لجميع مفردات المجتمع .

ويتوقع أن تكون العينة العنقودية أقل دقة من العينة العشوائية البسيطة وذلك لأن بيانات العناقيد قد تكون اكثر تجانساً من بيانات العينة العشوائية البسيطة . ويعتمد ذلك بالطبع على درجة التجانس بين مفردات كل عنقود ، اذ ترتفع كفاءة العينة كلما قلت درجة التجانس .

يمثل ما سبق عرضاً موجزاً لبعض أساليب المعاينة . وقد تستخدم هذه الأساليب أو أساليب أخرى اكثر تعقيداً عند اختيار العينات في الحياة العملية . وتمثل أساليب اختيار العينات وتحليل نتائجها جزءاً أساسياً من علم الاحصاء .

هـ ـ ملاحظات أخرى

١ ـ من المفيد قبل القيام بالعملية الفعلية لجمع البيانات ، أن يتم اجراء تجربة استطلاعية Pilot Survey or Pretest تجمع فيها البيانات من عينة صغيرة من مفردات الدراسة . وتستخدم نتائج هذه التجربة فيما يلي :

- اختيار جودة صحيفة البحث والكشف عن أي عيوب بها ، والعمل على تعديلها قبل البدء في جمع البيانات .
- اختبار كفاءة جهاز جمع البيانات وتقويم مستوى تدريب جامعي البيانات ودراسة أية مشكلات قد تسبب عدم تعاون مفردات الدراسة .
- الحصول على معلومات إضافية عن مفردات المجتمع يمكن استخدامها لتحقيق تصميم أفضل لعينة الدراسة .

٢ ـ قد يلاحظ بعد الانتهاء من جمع البيانات ، أن معدل استجابة مفردات الدراسة غير كاف لتحقيق أهداف الدراسة . ويجب أن يكون هناك أساليب متفق عليها لمعالجة مثل هذا الموقف تتعلق بكيفية الاتصال ومتابعة مفردات الدراسة لحثهم على الادلاء بالبيانات المطلوبة ، وكيفية تحديد مدى عدم تحيز البيانات التي تجمع فعلاً في تمثيل بيانات المجتمع المستهدف .

٣ ـ ينبغي مراجعة وتدقيق البيانات بعد أن يتم جمعها وذلك للتأكد من أن جميع الأسئلة قد أجيب عنها بشكل واضح ومتسق ولاكتشاف أية أخطاء أو أنماط غير متوقعة في البيانات ، ثم العمل على تصحيح هذه الأخطاء . ويتم بعد ذلك ترميز هذه البيانات باستبدال الإجابات النوعية برموز رقمية وذلك تمهيداً لاستخراج الجداول الاحصائية المختلفة ثم استخدام أساليب الإحصاء الوصفي والاستنتاج الإحصائي لتحليل البيانات .

٥ ـ الأخطاء في البيانات الإحصائية Errors in Statistical Data

تتعرض البيانات الإحصائية للعديد من الأخطاء ، وذلك على الرغم من العناية التي قد تبذل عند جمع هذه البيانات ، ذلك أن إعداد وتنفيذ عملية جمع البيانات غالباً ما تكون محكومة باعتبارها اقتصادية واجتماعية متعددة . ولما كانت نتائج التحليل الإحصائي تتوقف على مدى جودة البيانات المستخدمة ، فإنه من الضروري تقويم نوعية البيانات الإحصائية ودراسة مدى انتشار الأخطاء فيها قبل البدء في عمليات تحليل وتفسير هذه البيانات . وتتألف أخطاء البيانات الإحصائية من عدة أنواع هي :

أ ـ **أخطاء الشمول أو التغطية Errors of Coverage** وهي الأخطاء التي يترتب عليها اختلاف بين تركيبة مفردات المجتمع المستهدف وتركيبة المفردات التي تدرس فعلاً . فمثلاً :

١ ـ عند اجراء تعدادات السكان ، دائماً ما يسقط بعض مفردات المجتمع أثناء عملية الحصر .

٢ ـ قد يستخدم إطار خاطىء عند اختيار عينة عشوائية من المجتمع ، مثال ذلك استخدام دليل الهاتف كإطار عند اختيار عينة عشوائية من سكان المدينة لدراسة رأي السكان في نظام مقترح للمواصلات العامة .

٣ ـ يلاحظ وجود قصور في تسجيل المواليد والوفيات في المناطق الريفية والمناطق البعيدة عن العمران .

٤ ـ عند إجراء دراسة بأسلوب الاستبيان الذاتي ، دائماً ما يهمل بعض

مفردات العينة الاستجابة للدراسة .

ب ـ أخطاء المضمون Errors of Content ويقصد بذلك الأخطاء التي تنشأ من اختلاف المعلومات الخاصة بمفردة ما عن الواقع ، فمثلاً :

١ ـ قد يكون هناك خطأ في أجهزة القياس المستخدمة لجمع البيانات مثل استخدام ميزان غير دقيق لقياس وزن الأفراد .

٢ ـ مغالاة الأفراد في مستواهم التعليمي ، أو إعطائهم بيانات خاطئة عن عمرهم وذلك في البيانات التي تجمع ضمن تعداد السكان .

٣ ـ ما قد يحدث من أخطاء عند نقل البيانات الإحصائية من مصدر لآخر .

٤ ـ قد يخطىء جامعو البيانات في تسجيل المعلومات التي يحصلون عليها من مفردات الدراسة عن عمد أو غير عمد .

٥ ـ استخدام أساليب التقريب عند تسجيل البيانات الإحصائية .

ويؤدي وجود أخطاء الشمول وأخطاء المضمون إلى تحيز البيانات في تمثيل المجتمع المستهدف من الدراسة ، ولذلك تسمى هذه الأخطاء بأخطاء التحيز Bias . ويلاحظ أنه يمكن التقليل من حجم هذه الأخطاء من خلال التحضير لعملية جمع البيانات بعناية ، ثم تنفيذها بجدية وإحكام .

حـ ـ أخطاء المعاينة Sampling Errors وتسمى أيضاً الأخطاء العشوائية ، ويقصد بذلك ما يلاحظ من أن النتائج التي تؤدي إليها العينة نادراً ما تتفق مع النتائج الفعلية في المجتمع . بل إن النتائج التي تؤدي إليها عينة ما نادراً ما تتفق مع النتائج التي تؤدي إليها عينة أخرى من نفس المجتمع . وينشأ ذلك نتيجة عوامل الصدفة التي تتحكم في اختيار مفردات العينة .

وتجدر الإشارة إلى أن حجم هذه الأخطاء يتوقف على نوع العينة العشوائية المستخدمة ، كما أن هذه الأخطاء تتناقص كلما كبر حجم العينة . وتهتم أساليب الاستنتاج الاحصائي بكيفية قياس حجم هذه الأخطاء ، ومن ثم تحديد درجة الثقة عند تعميم نتائج العينة إلى المجتمع الذي سحبت منه .

يتضح من ذلك أن مستوى الدقة يختلف من مجموعة لأخرى من البيانات الاحصائية . ويجب أن يؤخذ ذلك في الاعتبار عند عرض ومعالجة هذه البيانات . فمثلاً :

١ - لا ينبغي عرض البيانات بشكل يعطي انطباعاً خادعاً عن مستوى الدقة فيها . فمثلاً لا يجب القول بأن عدد سكان المدينة يساوي ١,٦٢١,٢١٥ نسمة وإنما يفضل القول بأن عدد سكان المدينة يساوي ١,٦ مليون نسمة وذلك بسبب الطبيعة التقريبية لهذه البيانات والتي تنشأ عن الصعوبات المعروفة التي تواجه عملية عد السكان . كذلك لا يجب القول بأن قيمة مشاريع التشييد والبناء في دولة ما خلال العام الماضي قد بلغت ٤,٥٣٢,٤٣٧ درهما ، وإنما قد يكتفي بالقول بأن هذه القيمة قد بلغت ٤,٥ مليون درهم وذلك للتأكيد على الطبيعة التقريبية لهذا البيان الذي تم تجميعه من مصادر عديدة قد تختلف فيما بينها في كيفية حساب قيمة منشآتها . ويجب ، كقاعدة عامة ، ألا يزيد مستوى الدقة في عرض النتيجة النهائية للعمل الاحصائي عن أقل مستوى للدقة في البيانات المستخدمة في الحساب . فمثلاً إذا أظهرت دائرة المعارف أن مساحة أوروبا تساوي ٣,٧٦٩,١٠٧ ميلاً مربعاً وأن مساحة آسيا تساوي ١٧,٣ مليون ميل مربع فإن مساحة أوربا وآسيا معاً بناءاً على هذه البيانات يجب أن تكتب ٢١,١ مليون ميل مربع وليس ٢١,٠٦٩,١٠٧ ميلاً مربعاً .

٢ - يجب العمل على تقدير الحجم المتوقع للخطأ في البيان الإحصائي كلما كان ذلك ممكناً . مثال ذلك ما ورد في أحد التقارير من أن قيمة احتياطيات دولة ما من الذهب تبلغ عشرون بليوناً من الدولارات ، قد تزيد أو تنقص بمقدار عشرة ملايين من الدولارات وذلك نتيجة عدم إمكانية تحديد وزن الذهب بدرجة متناهية من الدقة. كذلك إذا قيل أن نسبة السكان الذين ينوون قضاء الصيف بالخارج تبلغ ٠,٣٨ بناءاً على نتائج عينة عشوائية من أسر المجتمع ، فإنه يجب حساب حجم خطأ المعاينة ودرجة الثقة في النتائج ، وبيان ما إذا كانت العينة قد تعرضت لأخطاء عدم الاستجابة .

تمرينات

١ - ما هو نوع المتغير الاحصائي في كل حالة من الحالات الآتية :

(أ) المساحة المنزرعة قمحاً في مناطق الدولة المختلفة .

(ب) درجات الحرارة اليومية المسجلة خلال شهر مايو من هذا العام .

(حـ) أوزان طلبة المدارس الثانوية بالكيلوجرامات .

(ء) عدد السيارات التي تمتلكها الأسرة في مجتمع معين .

(هـ) ترتيب الطلبة في امتحان الإحصاء الأخير .

(و) الدرجات التي حصل عليها الطلبة في امتحان الاحصاء الأخير .

(ز) توزيع سكان الدولة حسب الديانة .

٢ - هل المتغير الاحصائي متقطع أم متصل في كل حالة من الحالات الآتية :

(أ) عدد الطلبة المسجلين في الجامعة خلال السنوات الخمس الماضية .

(ب) متوسط ما تنفقه الأسرة في أحد المجتمعات سنوياً على الطعام والشراب.

(حـ) عدد حوادث القتل في مدن الدولة خلال عام ١٩٨٥ .

(ء) أوزان الأبقار في إحدى المزارع .

(هـ) قيمة الناتج الاجمالي المحلي لمجموعة من الدول خلال هذا العام .

٣ - فيما يلي مجموعة البيانات الخاصة بالأفراد المتقدمين لشغل وظيفة ما :

الإسم	العمر	النوع	عدد البرامج التدريبية التي اشترك فيها	ترتيب الأفراد في رأي لجنة التعيينات
علي	٤٥	ذكر	٢	٢
ابراهيم	٣٢	ذكر	١	٣
فاطمة	٤٣	انثى	٣	١

أ ـ ما هي مفردات الدراسة في هذه الحالة ؟

ب ـ ما هي المتغيرات المتضمنة في مجموعة البينات ؟ وما هي أنواع هذه المتغيرات ؟

حـ ـ ماهي المشاهدات الخاصة بالعمر ؟

٤ ـ استخدم مكتبة الجامعة للبحث عن مصادر للبيانات الآتية ، وبين في كل حالة ما إذا كان المصدر أولياً أم ثانوياً ثم اذكر اسم جهة النشر :

أ ـ عدد سكان الدولة في تعداد السكان الأخير .

ب ـ عدد المواليد خلال العام الماضي .

جـ ـ أعداد الطلبة والمدرسين في مراحل التعليم المختلفة بالدولة خلال العام الماضي .

ء ـ عدد حوادث المرور في مناطق الدولة المختلفة خلال العام الماضي .

هـ ـ عدد تصاريح العمل الممنوحة في الدولة خلال السنوات الخمس الماضية .

٥ ـ يراد استطلاع رأي أفراد الأسرة الجامعية حول اقتراح بفتح المكتبة أيام الجمع والعطلات الرسمية . ما هي الخطوات الواجب اتباعها في تصميم وتنفيذ هذه الدراسة .

٦ ـ تقوم إحدى شركات بيع المواد الغذائية ببيع منتجاتها معبأة في زجاجات في المدينة أ ، بينما تبيع نفس المنتجات معبأة في علب في المدينة ب . قام قسم الأبحاث في الشركة بتحليل بيانات المبيعات في المدينتين لدراسة مدى تفضيل المستهلكين لأسلوب التعبئة . هل هذه الدراسة تجريبية أم غير تجريبية ؟ ولماذا ؟

٧ ـ قسمت عينة من ٣٠ طالباً جامعياً إلى ١٥ مجموعة تحتوي كل منها على طالبين متشابهين بقدر الامكان حسب العمر والخبرة التعليمية والخلفية

العامة . اختير أحد طالبي كل مجموعة عشوائياً ليدرس اللغة الانجليزية بنظام جديد يعتمد على الأساليب السمعية والبصرية بينما يدرس الطالب الآخر تبعاً للنظام التقليدي . ثم أعطي لجميع الأفراد نفس الاختبار في نهاية الفصل الدراسي لقياس مدى تحصيل كل منهم .

أ ـ هل هذه الدراسة تجريبية أم لا ؟ ولماذا ؟

ب ـ اذا اقترح أحد الأشخاص اجراء تعديل في تصميم الدراسة بحيث يتفق طالبي كل مجموعة فيما بينهما على الطريقة التي يدرس بها كل منهم . ما رأيك في هذا الاقتراح ؟

٨ ـ اذكر في كل حالة من الحالات الآتية ما إذا كانت طريقة المقابلة المباشرة أو طريقة الاستبيان الذاتي أو طريقة الملاحظة أكثر ملاءمة لجمع البيانات . وضح سبب الإجابة .

أ ـ بيانات عن مستوى الكولسترول في الدم لسكان المدينة .

ب ـ بيانات عن الدرجة العلمية ومكان العمل الحالي للأعضاء المسجلين في الجمعية الاحصائية .

حـ ـ بيانات عن رأي المستهلكين في نوع جديد من الصابون .

ء ـ بيانات عن أنماط الانفاق على السلع والخدمات المختلفة لأسر المجتمع خلال شهر معين .

هـ ـ بيانات عن أنماط مشاهدة برامج التليفزيون بين أسر المدينة .

و ـ بيانات من أصحاب الأعمال عن عدد العمال الذين يشتغلون لديهم خلال شهور السنة المختلفة .

٩ ـ لماذا يعتبر كل من الأسئلة الآتية غير مناسب للحصول على البيانات المطلوبة ؟

أ ـ هل يعتبر دخل أسرتك مرتفعاً أو متوسطاً أو منخفضاً إذا ما قورن بمستوى دخول الأسر في الحي الذي تسكن فيه ؟

ب ـ كم عدد أنابيب معجون الأسنان التي اشتريتها خلال الأربع

والعشرين شهراً الماضية ؟

حـ ـ ألا تعتقد أن على طلبة الجامعة واجب وطني للمساهمة في مشروع محو الأمية ؟

ء ـ هل أصيب أحد من أعضاء أسرتك بارتفاع غير طبيعي في درجة الحرارة مؤخراً ؟

١٠ ـ اذكر نوع العينة المستخدم في كل حالة من الحالات الآتية :

أ ـ خلطت الأوراق في مجموعة لأوراق اللعب خلطاً جيداً ثم اختير منها عينة من ١٣ ورقة .

ب ـ في دراسة عن خصائص المساكن في مدينة ما ، اختيرت عينة عشوائية من أحياء المدينة ثم فحصت جميع مساكن هذه الأحياء .

حـ ـ في دراسة لتقدير نسبة المواطنين الذين ينوون قضاء الصيف بالخارج ، قسم المواطنون إلى ثلاث مجموعات حسب مستوى الدخل ، ثم سحبت عينة عشوائية من مائة أسرة من كل مجموعة .

ء ـ في دراسة عن رأي السكان في تغيير مواعيد دوام الموظفين ، قام الباحث بالوقوف على ناصية أحد الميادين العامة وجمع بيانات من أول خمسين شخصاً مروا عليه .

هـ ـ في دراسة تتعلق بشركات السياحة في الدولة ، استخدم الباحث أحدث دليل متاح لهذه الشركات وعمره خمس سنوات كإطار للمعاينة تم على أساسه اختيار عينة عشوائية بسيطة من ٥٠ شركة .

١١ ـ اذكر نوع الخطأ الموجود في كل من البيانات التالية :

أ ـ عند نقل البيانات المتعلقة بالمواليد في مدينة ما، قام الموظف المختص بتسجيل وزن طفل على أنه ٦,٣ كيلوجراماً بدلاً من وزنه الفعلي ٣,٦ كيلوجراماً ، عن غير عمد .

ب ـ عند جمع بيانات تعداد السكان ، يتهرب الأشخاص الذين ليس لديهم إقامة قانونية في الدولة من الادلاء بالبيانات .

حـ - عند جمع بيانات عن مهن أفراد المجتمع ، قام الأفراد بإعطاء مهنهم تبعاً لتخصصاتهم الدراسية وليس تبعاً لمواصفات الوظائف التي يشغلونها .

ء - عند دراسة آراء المحامين حول قانون معين ، جمعت بيانات من المحامين المقيمين في عاصمة الدولة فقط .

هـ - في دراسة عن آراء الطلبة حول التدخين ، أرسلت استبيانات إلى عينة عشوائية من ٢٠٠ طالباً ، قام ٧٠ طالباً منهم فقط بالاستجابة للدراسة .

١٢ - تقوم إحدى شركات الأدوية بإجراء دراسة بطريقة الاستبيان الذاتي بين الأطباء للتعرف على نسبة الأطباء الذين يصفون أحد العقاقير التي تنتجها هذه الشركة لمرضاهم . اختيرت عينة عشوائية من ٢٠٠٠ طبيب فوجد أن نسبة الأطباء في العينة الذين يصفون هذا العقار = ٠,٤٠ وذلك بافتراض أن معدل الاستجابة للدراسة يساوي ١٠٠٪ .

أ - افترض أن معدل الاستجابة انخفض إلى ٩٥٪ . اشرح كيف يمكن أن يؤثر ذلك على قيمة النسبة المحسوبة من العينة ؟

ب - كرر الجزء (أ) إذا انخفض معدل الاستجابة إلى ٢٠٪ . اشرح دلائل هذه النتائج من حيث تأثير معدل الاستجابة على درجة الدقة في نتائج الدراسة .

حـ - هل يمكن اعتبار المستجيبين في هذه الحالة عينة ممثلة للمجتمع ؟ وضح سبب إجابتك .

١٣ - يراد اختيار عينة عشوائية بسيطة من مائة طالب لدراسة ظاهرة التغيب عن حضور المحاضرات باستخدام طريق الاستبيان الذاتي . لاحظ الباحث من الخبرات السابقة أن معدل الاستجابة في مثل هذه الدراسات يساوي تقريباً ٢٥٪ ، لذلك اقترح اختيار عينة عشوائية من ٤٠٠ طالب وذلك حتى يضمن مساهمة مائة منهم في الدراسة . ما رأيك في سلامة هذا الاقتراح ؟ وضح سبب اجابتك .

١٤ ـ علق على مدى صحة العبارات التالية ، مع توضيح سبب الإجابة .

(أ) تعتبر بيانات تعداد السكان بيانات كاملة الدقة لأنها تجمع على أساس حصر شامل لجميع مفردات المجتمع .

(ب) تتساوى المصادر الأولية والمصادر الثانوية في درجة الدقة ومستوى التفصيل في البيانات .

(حـ) لا تستخدم التجربة الميدانية إلا نادراً وفي المواقف التي تسمح فيها الميزانية المتاحة بذلك .

(ء) يؤدي استخدام أسلوب العينات إلى الحصول على بيانات أكثر دقة وأعلى تفصيلاً من أسلوب الحصر الشامل .

(هـ) لا يمكن استخدام أساليب الاستنتاج الاحصائي لتعميم نتائج العينات غير الاحتمالية إلى المجتمعات التي تسحب منها .

١٥ ـ استخدمت عينة عشوائية من الأشخاص لدراسة تأثير التدخين على الصحة . اتضح من نتائج العينة أن الأشخاص الذين لم يعتادوا التدخين أكثر صحة بشكل عام من الأشخاص الذين يدخنون ، وأن الأشخاص الذين يدخنون أفضل صحة بشكل عام من الأشخاص الذين أقلعوا عن التدخين . كذلك وجد أن هذه النتائج تنطبق على جميع فئات الأعمار في العينة .

(أ) ما هي الحكمة في دراسة كل فئة من فئات الأعمار على حدة ؟

(ب) يدعي أحد الباحثين أن هذه النتائج تشير إلى أنه لا ينبغي للشخص أن يبدأ في التدخين ، كما لا يجب أن يقلع المدخنون حالياً عن التدخين . هل توافق على هذا الادعاء ؟ وضح سبب إجابتك .

١٦ ـ حدد بمجرد النظر ما إذا كانت النسب الآتية أقرب إلى ١٪ أو ١٠٪ أو ٢٥٪ أو ٥٠٪ .

(أ) ٣٩ إلى ٣٩٨   (ب) ٩٩ الى ٤٠٧

(حـ) ٥٧ إلى ٢٠٩   (ء) ٩٩ إلى ١٩٧

١٧ - في عينة من ٤٤٦ أسرة وجد ٤٦ أسرة يتراوح دخلها الشهري بين ٥ آلاف ، ١٠ آلاف درهم .

(أ) ما هي نسبة الأسر التي يتراوح دخلها بين خمسة آلاف وعشرة آلاف درهم ؟

(ب) قدر نسبة الأسر التي يتراوح دخلها بين سبعة آلاف وتسعة آلاف درهم ؟



الباب الثالث

التوزيعات التكرارية

١ ـ مقدمة

تختلف قيم المتغير الإحصائي من مفردة إلى أخرى بين المفردات المستهدفة في الدراسة . وتنبع أهمية الأساليب الاحصائية من ملاءمتها لوصف وتحليل أنماط هذه الاختلافات من خلال جمع بيانات عن المتغيرات الاحصائية ثم تحليلها .

تتألف الدراسة الإحصائية لنمط الاختلاف في متغير ما من محاولة استخدام مجموعة البيانات المتاحة للاجابة عن أسئله واضحة ومحددة تتعلق بالأمور التالية :

(أ) ما هو الشكل العام للاختلاف في قيم المتغير ؟

(ب) ما هي طبيعة العلاقة بين نمط الاختلاف في المتغير وأنماط الاختلاف في متغيرات أخرى ؟

(حـ) اذا كانت البيانات تمثل عينة ، كيف يمكن تعميم نتائج (أ) ، (ب) إلى المجتمع ككل .

فمثلاً ، إذا كانت لدينا مجموعة بيانات عن عينة من ٢٠٠ طالب من طلبة الجامعة، تحتوي على بيانات مشابهة لتلك التي تظهر في شكل (١) صفحة (٣٢) عن نوع الطالب وجنسيته وفرقته الدراسة ومعدله الدراسي العام وعدد أفراد أسرته والمسافة التي يقطعها يومياً في الذهاب إلى الجامعة ، فإن الدراسة الاحصائية لهذه المتغيرات تتطلب الإجابة عن أسئلة من النوع التالي:

١ ـ ما هو توزيع مفردات العينة حسب النوع ؟ أي ما هو عدد الذكور وما هو عدد الاناث بين مفردات العينة ؟

٢ ـ ما هو توزيع مفردات العينة حسب الجنسية ؟

٣ ـ ما هو توزيع مفردات العينة حسب الفرقة الدراسية ؟

٤ ـ كيف يختلف المعدل الدراسي العام بين طلبة العينة ؟ ما هو أكبر معدل وما هو أصغر معدل وما هو متوسط هذه المعدلات في العينة ؟

٥ ـ هل تختلف قيم المعدل الدراسي العام بين الذكور والاناث ؟

٦ ـ ما هو متوسط المسافة التي يقطعها الطالب يومياً في ذهابه إلى الجامعة ؟ ما هي أقصر هذه المسافات وما هي أطولها ؟

٧ ـ هل يختلف المعدل الدراسي العام للطالب باختلاف المسافة التي يقطعها يومياً في الذهاب إلى الجامعة ؟

٨ ـ هل يختلف المعدل الدراسي العام للطالب باختلاف عدد أفراد أسرته ؟

٩ ـ كيف يمكن تعميم الإجابات عن الأسئلة السابقة إلى مجتمع طلبة الجامعة عموماً ؟ .

تتطلب الإجابة عن مثل هذه الأسئلة البدء بتنظيم البيانات ووضعها في شكل جداول وتوزيعات تكرارية ، ثم استخدام ذلك كأساس للعمليات الاحصائية التالية . وينشأ التوزيع التكراري بتصنيف مفردات الدراسة تبعاً للقيم المختلفة للمتغير محل الاهتمام ، وتحديد عدد مرات تكرار كل قيمة أو مجموعة من قيم هذا المتغير .

وتجدر الإشارة إلى أن تنظيم وجدولة البيانات الاحصائية قد يكون هدفاً في حد ذاته . مثال ذلك ما تقوم به الجهات المختلفة لجمع البيانات الاحصائية من تنظيم البيانات التي تشرف على جمعها ثم جدولتها ونشرها في شكل مفيد . وقد ازدادت أهمية عمليات تنظيم وجدولة البيانات في الآونة الأخيرة نتيجة ما صاحب الاستخدام الواسع للأساليب الكمية في دراسة العلوم المختلفة من توافر كم هائل من البيانات الاحصائية عن كافة أوجه النشاط

الانسـاني . هذا بـالإضافـة إلى أن انتشار الحـاسبات الآليـة وتوافـر البرامـج المناسبة لتشغيلها قد سهل التعامل مع هذه البيانات وتنظيمها على نحو سريع وكفء .

ويعتمد أسلوب انشاء الجداول والتوزيعات التكرارية على نوع البيانات المستخدمة . اذ يلاحظ مثلاً أن البيانات النوعية تتطلب أساليب أكثر بساطة من تلك المستخدمة مع البيانات الكمية ، كما أن طرق تحليل البيانات المتقطعة تكـون أقل تعقيـداً من طرق تحليـل البيـانـات المتصلة . وفيمـا يلي عـرض للأساليب المختلفة لتنظيم البيانات وإنشاء التوزيعات التكرارية .

٢ ـ التوزيع التكراري للبيانات النوعية

يتكون جدول التوزيع التكراري من عمودين ، يعطي العمود الأول قائمة بالأوجه المختلفة للمتغير محـل الدراسـة بينما يتم في العمـود الثاني تصنيف مفردات الدراسة على تلك الأوجه . ولعل أبسط الأمثلة عن المتغيرات النوعية هو نوع الفرد (ذكر/ أنثى) . يعطي جدول (١) بيانات جمعت عن نوع الفرد في عينة حجمها ٢٠ من موظفي وزارة التربية ، حيث يلاحظ أن نوع الفرد هو المتغيـر محل الاهتمـام وأن الموظفين يمثلون مفـردات الـدراسـة وأن هنـاك عشرون مشاهدة ، واحدة لكل موظف .

جدول (١)
بيـانات عن نـوع الشخص في عينة افتـراضية
حجمهـا ٢٠ من موظفي وزارة التربية

(١) ذكر	(٥) ذكر	(٩) أنثى	(١٣) ذكر	(١٧) أنثى
(٢) أنثى	(٦) ذكر	(١٠) أنثى	(١٤) ذكر	(١٨) ذكر
(٣) ذكر	(٧) أنثى	(١١) ذكر	(١٥) أنثى	(١٩) أنثى
(٤) ذكر	(٨) ذكر	(١٢) أنثى	(١٦) ذكر	(٢٠) ذكر

إذا كان الهدف هو التعرف على توزيع مفردات العينة حسب النوع ، فإن هذا الشكل الخام للبيانات لا يفيد كثيراً ، ويكون من الضـروري تنظيم هـذه البيانات ووضعها في شكل جدول توزيع تكراري . وينشأ هذا الجدول ، كما سبق القول ، بتحديد الأوجه المختلفة للمتغير (ذكر، أنثى في هذه الحالة) ثم تصنيف المشاهدات على هذه الأوجه . وتظهر نتيجة هذه العملية في جدول (٢) .

جدول (٢)
التـوزيع التكـراري لمفردات عينـة حجمها ٢٠
من مـوظفي وزارة التربيـة ، ١٩٨٥ ، حسب النوع

النوع	عدد الموظفين
ذكور	١٢
إناث	٨
المجموع	٢٠

(المصدر : جدول (١))

ويطلق على جدول (٢)إسم «توزيع تكـراري» Frequency Distribution أو بـاختصـار «تـوزيـع » . إذ أن هـذا الجـدول يبين كيفـيـة تـوزيـع مفردات العينة على الأوجه المختلفة للمتغير ، أو بعبارة أخرى ، عدد المرات التي يتكرر فيهاظهور كل وجه من هذه الأوجه بين مفردات العينة . ويلاحظ أن الانتقال من البيانات الخام في جدول (١) إلى التوزيـع التكراري في جـدول (٢) يتضمن نوعاً من التجـريد والتبسيط يتمثـل في التركيـز على نوع الأفـراد فقط ، دون الالتفات إلى أية تفـاصيل أخـرى تتعلق بهويـات هؤلاء الأفراد . فمثلاً يتضح من الجدول أن هناك ثمان إناث في العينة ، لم يعد من الممكن التعرف على شخصياتهن دون الرجوع إلى مجموعة البيانات الأصلية في جدول (١) . ويمكن أن يؤخذ ذلك في الاعتبار عند تحديد عنـوان الجدول بـالنص

على أن الجدول يعطي « التوزيع النوعي في العينة » .

يوضح جدول (٢) الخصائص المطلوبة في الجدول الاحصائي الجيد . هذه الخصائص هي :

(أ) **البساطة** : يجب أن يكون الجدول بسيطاً بقدر الإمكان . وفي هذا الصدد ينصح بأن يستخدم كل جدول لتمثيل البيانات اللازمة للإجابة عن سؤال محدد واحد فقط .

(ب) **عنوان الجدول**: يجب أن يكون للجدول عنواناً واضحاً وكاملاً يشمل نوعية البيانات التي تظهر في الجدول ومكان الحصول على هذه البيانات بالإضافة إلى بعدها الزمني .

(حـ) **عناوين الأعمدة والصفوف في الجدول** : يجب أن يكون كل عمود وكل صف في الجدول معنوناً بشكل واضح وموجز . اذا كانت هذه العناوين طويلة وتزيد عن المساحات المخصصة لها ، يمكن الاكتفاء بملخص لكل عنوان ثم إعطاء العناوين الكاملة في حواشي خاصة .

(ء) **خلايا الجدول** : يتألف الجدول من عدة خلايا . وتنشأ الخلية من تقاطع أحد أعمدة الجدول مع أحد صفوفه . ويلاحظ أن التوزيع التكراري يتكون بتصنيف المشاهدات على هذه الخانات .

(هـ) **مصدر البيانات** : يجب دائماً أن يذكر مصدر البيانات الخام التي استخدمت في إنشاء الجدول .

ويمكن تحويل التوزيع التكراري إلى توزيع تكراري نسبي بقسمة التكرارات المختلفة على العدد الكلي للمفردات . وتمثل التكرارات النسبية نسب تكرار الأوجه المختلفة للمتغير في العينة . ويمكن كذلك ضرب هذه النسب في ١٠٠ للحصول على تكرارات نسبية مئوية . ويعطي جدول (٣) التوزيع النوعي النسبي لمفردات العينة المناظر للتوزيع التكراري في جدول (٢) . ويلاحظ ضرورة أن يكون مجموع النسب دائماً هو الواحد الصحيح (فيما عدا الاختلافات البسيطة التي قد تنشأ من تقريب الأرقام) .

ويلعب مفهوم النسبة دوراً هاماً في تفسير البيانات النوعية ، فمثلاً تقاس ظاهرة البطالة بحساب نسبة العاطلين إلى قوة العمل وتقاس ظاهرة النوع بحساب نسبة الذكور إلى مجموع السكان وهكذا . ويجب تفادي استخدام النسب اذا كان المقام المستخدم لحسابها صغيراً لأن أي تغير طفيف في قيمة البسط في هذه الحالة قد ينعكس في تغير قيمة النسبة تغيراً كبيراً . ولذلك يجب عند استخدام النسب في التحليل أن تعطي أيضاً البيانات الأصلية التي حسبت منها هذه النسب ، حتى يتمكن القارىء من تفسير النتائج بشكل سليم .

جدول (٣)

التوزيع النوعي النسبي في عينة حجمها ٢٠ من موظفي وزارة التربية

النوع	نسبة الموظفين (التكرار النسبي)	النسبة المئوية
ذكور	$\frac{١٢}{٢٠}$ = ٠,٦	٦٠
اناث	$\frac{٨}{٢٠}$ = ٠,٤	٤٠
المجموع	١,٠	١٠٠

(المصدر : جدول (٢))

ويستفاد من التوزيع التكراري النسبي في دراسة الأهمية النسبية لأوجه المتغير المختلفة بالإضافة إلى استخدامه كأساس لإجراء المقارنات بين عدد من التوزيعات التكرارية . فمثلاً يعطي جدول (٤) التوزيع المهني للأشخاص في قوة العمل في إحدى الدول في عامي ١٩٦٠ ، ١٩٨٠ :

جدول (٤)

التوزيع المهني للأشخاص في قوة العمل في بلد ما في عامي ١٩٦٠ ، ١٩٨٠

المهنة	عدد الأشخاص	
	١٩٨٠	١٩٦٠
مهن مكتبية	٣٨٠٦٨	٢٢٣٧٣
مهن يدوية	٢٧٤٥٢	٢٣٣٣٦
عمال الخدمات	٩٧٢٤	٦٥٣٥
العمال الزراعيون	٣١٦٤	٧٤٠٨
المجموع	**٧٨٤٠٨**	**٥٩٦٥٢**

(المصدر : بيانات افتراضية)

إذا كان الهدف هو استخدام هذه البيانات لدراسة التغيرات التي حدثت في التركيبة المهنية لقوة العمل بين عامي ١٩٦٠، ١٩٨٠، فإن التوزيعات النسبية تكون اكثر وضوحاً وسهولة في الاستخدام، وتظهر هذه التوزيعات في جدول (٥) .

جدول (٥)

التوزيع المهني النسبي لقوة العمل في بلد ما في عامي ١٩٦٠ ، ١٩٨٠

المهنة	التكرار النسبي	
	١٩٨٠	١٩٦٠
مهن مكتبية	٠,٤٩	٠,٣٨
مهن يدوية	٠,٣٥	٠,٣٩
عمال الخدمات	٠,١٢	٠,١١
العمال الزراعيون	٠,٠٤	٠,١٢
المجموع	١,٠٠	١,٠٠
مجموع الأشخاص في قوة العمل	**٧٨٤٠٨**	**٥٩٦٥٢**

(المصدر : جدول (٤))

ويتضح من هذا الجدول أن نسبة العاملين في المهن المكتبية قد ارتفعت ارتفاعاً واضحاً بين عامي ١٩٦٠ ، ١٩٨٠ على حين انخفضت نسبة العمال الزراعيين خلال نفس الفترة . كذلك يلاحظ أن نسبة العاملين في المهن اليدوية قد انخفضت انخفاضاً طفيفاً بينما ارتفعت نسبة العاملين في الخدمات ارتفاعاً طفيفاً . وقد ترتب على ذلك أن أصبحت المهن المكتبية اكثر المهن أهمية في التوزيع التكراري لعام ١٩٨٠ .

تستخدم نفس المبادىء السابقة عند انشاء التوزيعات التكرارية للبيانات الترتيبية . يعطي جدول (٦) المشاهدات التي جمعت من ٤٠ سائقاً ، سئل كل منهم عن تقييمه لدرجة سلاسة القيادة في سيارته :

جدول (٦)

درجة سلاسة قيادة السيارة في عينة افتراضية من ٤٠ سائقاً

(١) ممتاز	(٥) سيء	(٩) جيد	(١٣) جيد	(١٧) مقبول
(٢) جيد جداً	(٦) سيء جداً	(١٠) جيد جداً	(١٤) جيد	(١٨) جيد
(٣) جيد	(٧) جيد	(١١) جيد	(١٥) جيد جداً	(١٩) جيد
(٤) مقبول	(٨) جيد	(١٢) جيد	(١٦) جيد	(٢٠) جيد جداً
(٢١) مقبول	(٢٥) جيد جداً	(٢٩) مقبول	(٣٣) جيد جداً	(٣٧) جيد
(٢٢) جيد	(٢٦) جيد	(٣٠) مقبول	(٣٤) ممتاز	(٣٨) جيد
(٢٣) جيد	(٢٧) جيد	(٣١) جيد جداً	(٣٥) جيد جداً	(٣٩) جيد جداً
(٢٤) ممتاز	(٢٨) جيد	(٣٢) جيد	(٣٦) مقبول	(٤٠) سيء

ويظهر التوزيع التكراري المناظر لهذه البيانات الترتيبية في جدول (٧) .

جدول (٧)

التوزيع التكراري لدرجة سلاسة القيادة في عينة افتراضية من ٤٠ سائقاً

درجة السلاسة	عدد السائقين	التكرار النسبي
ممتاز	٣	٠,٠٧٥
جيد جداً	٩	٠,٢٢٥
جيد	١٩	٠,٤٧٥
مقبول	٦	٠,١٥٠
سيء	٢	٠,٠٥٠
سيء جداً	١	٠,٠٢٥
المجموع	٤٠	١,٠٠٠

(المصدر : جدول (٦)) .

ويلاحظ أن العلاقة الترتيبية بين أوجه المتغير في جدول (٧) تسمح بتجميع التكرارات بشكل مفيد . فمثلاً يمكن حساب نسبة السائقين الراضين عن سلاسة سياراتهم عموماً بتجميع التكرارات النسبية المناظرة للأوجه ممتاز وجيد جداً وجيد لنحصل على ٠٧٥, + ٢٢٥, ٠ + ٤٧٥, ٠ = ٧٧٥, ٠ كذلك فإن نسبة السائقين المستائين من سلاسة سياراتهم تساوي ٠٥٠, + ٠٢٥, = ٠٧٥, وهي مجموع النسب المناظرة للأوجه سيء وسيء جداً ، وهكذا . ويوضح ذلك قابلية البيانات الترتيبية للاستخدام العددي بشكل يفوق قابلية البيانات التصنيفية لهذا الاستخدام .

٣ ـ التوزيع التكراري للمتغيرات الكمية المتقطعة

يتم إنشاء الجدول التكراري في هذه الحالة بالطريقة المعتادة ، حيث تظهر القيم الممكنة للمتغير في أحد أعمدة الجدول بينما يظهر في العمود الآخر عدد مرات حدوث أو تكرار كل قيمة من هذه القيم بين مفردات

الدراسة . كذلك يمكن الحصول على التكرارات النسبية بالقسمة على العدد الكلي للمفردات . فمثلاً ، يعطي جدول (٨) عدد الأطفال في الأسرة لعينة افتراضية من ٣٠ أسرة ، حيث يلاحظ أن مفردة الدراسة هي الأسرة وأن المتغير محل الاهتمام هو عدد أطفال الأسرة وأن هناك ثلاثون مشاهدة ، واحدة لكل أسرة .

جدول (٨)

عدد أطفال الأسرة في عينة افتراضية من ٣٠ أسرة

(١)	صفر	(٧)	٢	(١٣)	٣	(١٩)	٢	(٢٥)	١
(٢)	١	(٨)	٤	(١٤)	٢	(٢٠)	١	(٢٦)	٤
(٣)	٢	(٩)	١	(١٥)	٣	(٢١)	٣	(٢٧)	٢
(٤)	٣	(١٠)	٢	(١٦)	١	(٢٢)	صفر	(٢٨)	٣
(٥)	٢	(١١)	صفر	(١٧)	٤	(٢٣)	٤	(٢٩)	صفر
(٦)	٣	(١٢)	٣	(١٨)	٢	(٢٤)	٢	(٣٠)	٢

ويعطي جدول (٩) التوزيع التكراري المناظر لهذه البيانات .

ويلاحظ أن الطبيعة الكمية للمتغير محل الدراسة تمكن من حساب مقاييس عددية مختلفة من جدول التوزيع التكراري . فمثلاً إذا أردنا حساب نسبة الأسر التي يقل عدد أفرادها عن ٣ ، تجمع التكرارات النسبية المناظرة ١٣٣ , + ٠,١٦٧ + ٠,٣٣٣ = ٠,٥٣٣ ، كما أن عدد الأسر التي لها ثلاثة أطفال على الأقل يساوي ٧ + ٤ = ١١ أسرة وهكذا . وسنرى فيما بعد كيفية حساب مقاييس عددية أخرى مثل متوسط عدد الأطفال للأسرة وغيره من المقاييس التي تصف الخصائص العامة لنمط الاختلاف في المتغير . ويدل ذلك على أن الأساليب الاحصائية المستخدمة لتحليل هذه البيانات تكون أكثر تعقيداً من تلك المستخدمة مع البيانات النوعية .

جدول (٩)

التوزيع التكراري لعدد أطفال الأسرة في عينة افتراضية من ٣٠ أسرة .

عدد أطفال الأسرة	عدد الأسر	التكرار النسبي	التكرار النسبي المئوي
صفر	٤	$\frac{٤}{٣٠}$ = ٠,١٣٣	١٣,٣
١	٥	$\frac{٥}{٣٠}$ = ٠,١٦٧	١٦,٧
٢	١٠	$\frac{١٠}{٣٠}$ = ٠,٣٣٣	٣٣,٣
٣	٧	$\frac{٧}{٣٠}$ = ٠,٢٣٣	٢٣,٣
٤	٤	$\frac{٤}{٣٠}$ = ٠,١٣٣	١٣,٣
المجموع	٣٠	١,٠٠٠	١٠٠

(المصدر : جدول (٨))

تستخدم التوزيعات التكرارية كذلك كأساس لإجراء المقارنات بين مجموعات البيانات المختلفة ، فمثلاً يمكن تحديد ما إذا كانت قيم إحدى المجموعات تزيد بشكل عام عن قيم المجموعة الأخرى ، أو العكس . في دراسة للمقارنة بين عدد الحوادث التي تقع لسائقي سيارات الأجرة في مدينتين أ ، ب ، أخذت عينة عشوائية حجمها ٧٠٠ من سائقي سيارات الأجرة في المدينة أ وعينة عشوائية أخرى حجمها ٥٠٠ من سائقي المدينة ب وسجل عدد الحوادث التي وقعت لكل سائق خلال السنوات الأربع السابقة . في هذه الحالة ، هناك ٧٠٠ مشاهدة لسائقي المدينة أ بالاضافة إلى ٥٠٠ مشاهدة لسائقي المدينة ب . نظمت هذه البيانات في شكل توزيعات تكرارية كما يظهر في جدول (١٠) .

جدول (١٠)

التوزيعات التكرارية لعدد الحوادث لسائقي المدينة أ وسائقي المدينة ب

عدد الحوادث	عدد السائقين		التكرار النسبي	
	المدينة أ	المدينة ب	المدينة أ	المدينة ب
صفر	١١٧	٣٠٠	,١٦٧	,٦٠٠
١	١٥٧	١٥٠	,٢٢٤	,٣٠٠
٢	١٥٨	٢٥	,٢٢٦	,٠٥٠
٣	١٠٧	٢٠	,١٥٣	,٠٤٠
٤	٧٨	صفر	,١١١	,٠٠٠
٥	٤٤	٣	,٠٦٣	,٠٠٦
٦	٢١	صفر	,٠٣٠	,٠٠٠
٧	٨	٢	,٠١١	,٠٠٤
٨	٦	صفر	,٠٠٩	صفر
٩	٤	صفر	,٠٠٦	صفر
المجموع	٧٠٠	٥٠٠	١,٠٠٠	١,٠٠٠

(المصدر : بيانات افتراضية)

وتظهر المقارنة في جدول (١٠) أن عدد الحوادث التي تقع لسائقي المدينة أ يزيد عموماً عن العدد الذي يقع لسائقي المدينة ب . إذ يلاحظ مثلاً أن ٦٠٪ من سائقي المدينة ب لم تقع لهم أية حوادث وأن ٩٠٪ منهم قد وقعت لهم حادثة واحدة أو أكثر ، وهي نسب تزيد كثيراً عن النسب المناظرة لسائقي المدينة أ الذين تصل عدد الحوادث التي تقع لكل منهم في بعض الأحيان إلى تسعة . وسنرى فيما بعد كيف تستخدم هذه التوزيعات التكرارية لحساب مقاييس عددية إضافية يمكن الاعتماد عليها أيضاً في أغراض المقارنة .

٤ ـ ترتيب البيانات الكمية

هناك أساليب متعددة لتنظيم وعرض البيانات الكمية بهدف التعرف على خصائص نمط الاختلاف في هذه البيانات . وقد تستخدم بعض هذه الأساليب كبديل عن إنشاء التوزيع التكراري ، وقد تستخدم بعضها كخطوة تمهيدية عند انشاء هذا التوزيع .

ويعتبر الترتيب التصاعدي أو التنازلي للمشاهدات Ordering أحد الأساليب المفيدة لتنظيم وعرض البيانات الكمية ، إذ يمكن ذلك من التعرف على أصغر قيمة واكبر قيمة في البيانات بالإضافة إلى دراسة نمط توزيع المشاهدات داخل هذا المدى . فمثلاً ، يعطي جدول (١١) عدد نبضات القلب بعد الانتهاء من أداء بعض التمارين الرياضية لمجموعة من ٣٠ طالباً :

جدول (١١)

عدد نبضات القلب لثلاثين طالباً

٨٢	٩٥	٩٢	٦٢	٨٥	٩٢
٨٢	٩٥	٧٠	٨٥	٨٤	٩٥
٩١	٨٢	٩٤	٧٦	٨٨	٩١
٨٧	٨٠	٦٨	٥٨	٧٦	٨٥
١١٠	٦٠	٧٥	٨٨	٦٤	٧٤

ويظهر الترتيب التصاعدي لهذه البيانات في جدول (١٢) ، حيث يلاحظ أن ٥٨ هو أصغر عدد نبضات القلب وأن ١١٠ هو أكبر عدد لها كما أن متوسط البيانات يقع بين الأرقام ٨٢ ، ٨٤ ، ٨٥ . ويلاحظ كذلك أن الغالبية العظمى للمشاهدات تقع داخل المدى (٨٠ ـ ٩٥) وأن هناك فجوة واسعة بين القيمتين الأخيرتين ٩٥ ، ١١٠ مما قد يشير إلى أن القيمة ١١٠ تتميز عن بقية المشاهدات أو تشذ عنها لسبب أو لآخر .

جدول (١٢)

الترتيب التصاعدي لعدد نبضات قلب ثلاثون طالباً

٥٨	٧٠	٨٠	٨٥	٨٨	٩٤
٦٠	٧٤	٨٢	٨٥	٩١	٩٥
٦٢	٧٥	٨٢	٨٥	٩١	٩٥
٦٤	٧٦	٨٢	٨٧	٩٢	٩٥
٦٨	٧٦	٨٤	٨٨	٩٢	١١٠

(المصدر : جدول (١١)) .

وتصبح المعلومات المتضمنة في هذه البيانات المرتبة أكثر وضوحاً إذا ما رسمت هذه البيانات كما يتضح من شكل (١) .



شكل (١) : عدد نبضات القلب لثلاثين طالباً

(المصدر : جدول (١٢)

٥ ـ عرض البيانات الكمية في شكل أغصان وأوراق

انتشـر استخـدام شكــل الأغصـان والأوراق Stem - and Leaf Display لعرض البيانات الاحصائية في الآونة الأخيرة نتيجة إمكانية الحصول على هذا الشكل بسهولة باستخدام الحاسبات الآلية . تقسم كل مشاهدة أولاً إلى جزءين يمثل أحدها غصناً بينما يمثل الجـزء الآخر ورقـة على هذا الغصن . وتختـار الأغصان بالطبع بحيث تسمح بوجود أكثر من ورقة على أي منها . فمثلاً عند

عرض بيانات دقات القلب في جدول (١١) ، يمكن أن يؤخذ رقم العشـرات في كل مشاهدة ليمثل الغصن وأن يؤخذ رقم الآحاد ليمثل ورقة على الغصن. وبالتالي فإن مشاهدة مثل ٩٥ يكون غصنها ٩ وورقتها ٥ . وينشأ شكل الغصن والأوراق بكتابة الأغصان في قائمـة عموديـة ثم كتابـة الأوراق المناظـرة لكل غصن أفقياً بجانب هذه الأغصان . وفيما يلي بعض الأمثلة .

يعطي جدول (١٣) الأجر اليومي بالدرهم لمفـردات عينة حجمهـا ٥٠ عاملاً من عمال الخدمات بالمدينة .

جدول (١٣)

الأجر اليومي بالدرهم لخمسين عاملاً من عمال الخدمات

٨٥	٩٥	١٢١	١٣٠	٩٥
١٥٢	١١٦	١١٩	٨٩	٨٨
١٠٢	٩٢	١١٠	١٢٣	١١٣
٨٣	١٢١	١١٠	٩٠	٨٧
١٢٥	١٠١	٧٧	١٠٤	٩٤
١١٧	١٠٥	١٠١	١٠٧	١١٠
١٣٦	١١٤	٩٧	١٠٧	١١٦
١١٧	٨٨	١٠٤	١٠٣	٨٨
١٤٣	١٠٠	١١٧	١٠٩	١٠٤
١٠٨	٩٠	١٠٧	١٠٧	١١٩

تتراوح المشاهدات بين ٧٧ ، ١٥٢ وبالتـالي فإنـه يمكن اختيار أرقـام العشرات ٧ ، ٨ ، ٩ ، ١٠ ، ١١ ، ١٢ ، ١٣ ، ١٤ ، ١٥ ، لتمثل الأغصان في هذه الحالة. تكتب هذه الأغصان في قائمـة عموديـة ثم توضـع أوراق كل غصن إلى جانبه. فمثلاً تعرض القيمة الأولى في الجدول أي ٨٥ بوضع ٥ إلى

جانب الغصن ٨ بينما تعرض القيمة التالية ١٥٢ بوضع ٢ إلى جانب الغصن ١٥ وتعرض القيمة الثالثة ١٠٢ بوضع ٢ الى جانب الغصن ١٠ وهكذا . ويظهر شكل الأغصان والأوراق المناظر لهذه البيانات في شكل (٢) . ويلاحظ أن هذا الشكل يوضح النمط العام للاختلاف في الأجور اليومية للعمال . إذ تتراوح الأجور بين ٧٧ ، ١٥٢ درهماً بينما تتركز معظم المشاهدات في الأغصان ٨ ، ٩ ، ١٠ ، ١١ . يلاحظ كذلك أن المتوسط العام للأجور يقع داخل الغصن ١٠ وأن التوزيع الاجمالي للأجور حول هذا الغصن يبدو متماثلاً بشكل تقريبي . تمثل المشاهدة ٧٧ أجراً متدنياً ، بينما تمثل القيم ١٤٣ ، ١٥٢ أجوراً مرتفعة بالنسبة للمستوى العام السائد للأجور في العينة .

الغصن	الأوراق
٧	٧
٨	٥ ، ٣ ، ٨ ، ٩ ، ٨ ، ٧ ، ٨
٩	٥ ، ٢ ، ٠ ، ٧ ، ٠ ، ٥ ، ٤
١٠	٢ ، ٨ ، ١ ، ٥ ، ٠ ، ١ ، ٤ ، ٧ ، ٤ ، ٧ ، ٧ ، ٣ ، ٩ ، ٧ ، ٤
١١	٧ ، ٧ ، ٦ ، ٤ ، ٩ ، ٠ ، ٠ ، ٧ ، ٣ ، ٠ ، ٦ ، ٩
١٢	٥ ، ١ ، ١ ، ٣
١٣	٦ ، ٠
١٤	٣
١٥	٢

شكل (٢) : الأغصان والأوراق لبيانات أجور ٥٠ عاملاً من عمال الخدمات
(المصدر : جدول ١٣)

يجب أن يكون عدد الأغصان المستخدمة كافياً لتوضيح النمط العام في البيانات . ذلك أن الاعتماد على عدد قليل من هذه الأغصان قد يترتب عليه ازدحامها بالأوراق بحيث يصبح من الصعب دراسة نمط الاختلاف داخل كل غصن . ومن ناحية أخرى ، لا ينبغي أن يكون عدد الأغصان كبيراً جداً حتى لا تكون هذه الأغصان هزيلة وخالية من الأوراق . كمثال على ذلك ، يوضح شكل (٣) الأغصان والأوراق المناظرة لبيانات جمعت من ٥٠ قرية من قرى الدولة عن عدد أشجار النخيل في كل قرية ، حيث لوحظ أن المشاهدات

تتراوح بين ١١٠ ، ٥٨٧ وبالتالي تقرر استخدام رقم خانة المئات لتمثيل الأغصان .

الغصن	الأوراق
١	٨٠،٣٩،٩٩،١٠،٨٠،٧٩،٧٥،٥٠،٩٩،٥٠،٥٠،٣٩،٥٠،٧٠،٨٠،٣٠،٩٠
٢	٤٨،٦٨،٣٩،٩٩،٩٠،٠٠،٥٥،٣٩،٠٠،٨٥،٦٠،٥٥،٤٩،٩٠،٩٨،٩٨
٣	١٩، ٨٠، ٩٢، ٢٨، ٧٥، ٢٠، ٩٩، ٩٩
٤	٤٨، ٣٩، ٨٨
٥	٨٧، ٣٠، ٣٠، ٤٩، ٢٠

شكل ٣ : الأغصان والأوراق لعدد أشجار النخيل في ٥٠ قرية

يلاحظ ازدحام الغصنين ١ ، ٢ نتيجة تركز عدد كبير من المشاهدات داخل المدى ١٠٠ ـ ٢٩٩ . ويترتب على ذلك صعوبة دراسة شكل توزيع هذه المشاهدات داخل الأغصان . ويمكن التغلب على ذلك بزيادة عدد الأغصان المستخدمة فمثلاً يمكن استبدال كل غصن بغصنين أحدهما للقيم (٠٠ ـ ٤٩) والآخر للقيم (٥٠ ـ ٩٩) . ويظهر شكل (٤) الأغصان والأوراق لنفس البيانات ، في هذه الحالة .

الأغصان	الأوراق
١'	٣٩ ، ١٠ ، ٣٩ ، ٣٠
١''	٨٠ ، ٩٩ ، ٨٠ ، ٧٩ ، ٧٥ ، ٥٠ ، ٩٩ ، ٥٠ ، ٥٠ ، ٧٠ ، ٨٠ ، ٩٠
٢'	٤٨ ، ٣٩ ، ٠٠ ، ٣٩ ، ٠٠ ، ٤٩
٢''	٦٨ ، ٩٩ ، ٩٠ ، ٥٥ ، ٨٥ ، ٦٠ ، ٥٥ ، ٩٠ ، ٩٨ ، ٩٨
٣'	١٩ ، ٢٨ ، ٢٠
٣''	٨٠ ، ٩٢ ، ٧٥ ، ٩٩ ، ٩٩
٤'	٤٨، ٣٩
٤''	٨٨
٥'	٣٠ ، ٣٠ ، ٤٩ ، ٢٠
٥''	٨٧

شكل (٤) : الأغصان والأوراق لعدد أشجار النخيل في ٥٠ قرية

(' تشير الى الأغصان التي تحمل الأوراق (٠٠ ـ ٤٩) بينما تشير '' الى الأغصان التي تحمل الأوراق (٥٠ ـ ٩٩)) .

ويلاحظ من شكل (٤) أن شكل التوزيع يميل إلى الالتواء في اتجاه اليمين وأن مركزه يقع داخل الغصن ٢ً وأن هناك قمتان للتوزيع إحداها عند الغصن ١ً والأخرى عند الغصن ٢ً .

قد يقترح البعض في المثال السابق استخدام رقمي خانات العشرات والمئات معاً لتمثيل الأغصان واستخدام رقم خانة الآحاد للأوراق . في هذه الحالة يكون الشكل الناتج قليل الفائدة لأن عدد الأغصان يكون كبيراً جداً ويكون معظمها خالياً من الأوراق . ويترك للقارىء انشاء هذا الشكل كتمرين .

ويمكن الاعتماد على شكل الأغصان والأوراق للمقارنة بين نمط الاختلاف من مجموعتين للبيانات . ويتم ذلك بوضع الأوراق الخاصة بمجموعة ما على الجانب الأيمن للأغصان بينما توضع الأوراق الخاصة بالمجموعة الأخرى على الجانب الأيسر ، كما يتضح من المثال التالي .

في دراسة عن تأثير الظروف الاقتصادية والاجتماعية للسكان على أنماط تغذيتهم ، أخذت عينة من ٤٤ شخصاً موسراً من سكان المدن وعينة أخرى من ٤٩ شخصاً فقيراً من سكان الريف وتم قياس مقدار الكولسترول في دم كل منهم (مقاساً بالميليجرامات لكل لتر) . يعطي شكل (٥) شكل الأغصان والأوراق الخاصة بهذه البيانات .

أوراق عينة سكان الريف	الأغصان	أوراق عينة سكان الحضر
٥	٩	
٨ ، ٨	١٠	
٤ ، ٥	١١	
٤ ، ٩ ، ٩	١٢	
٩ ، ١ ، ٦ ، ٦ ، ١ ، ٥	١٣	٤ ، ٣
٢ ، ٤ ، ٣، ٨، ٣، ٢، ٥، ٤، ٦، ٠	١٤	
٨ ، ٥ ، ٢ ، ٧ ، ٨ ، ٢	١٥	٥
٢ ، ٥ ، ٦	١٦	
١ ، ٢ ، ٣ ، ٤ ، ٥	١٧	٠ ، ٥ ، ٩
١ ، ٩ ، ٠	١٨	٩ ، ٨ ، ٤ ، ١
٧ ، ٤ ، ٢	١٩	٦ ، ٠ ، ٧ ، ٩
٤	٢٠	٦ ، ٠ ، ١ ، ٠ ، ٥ ، ٥ ، ٤ ، ٥ ، ١
	٢١	٧ ، ٤
٠ ، ٦ ، ٣	٢٢	٢ ، ٧ ، ٨ ، ٧ ، ٢
١	٢٣	٩ ، ٤ ، ٦ ، ٤
	٢٤	١ ، ٤ ، ٩ ، ٢
	٢٥	٢
	٢٦	
	٢٧	٣ ، ٩
	٢٨	٤ ، ٤ ، ٤

شكل (٥) : الأغصان والأوراق لمقدار الكولسترول في الدم (ملجرام/ لتر) لمفردات عينة من سكان الريف وأخرى من سكان الحضر .

يلاحظ من هذا الشكل أن هناك درجة اختلاف واسعة في مشاهدات كل مجموعة ، إذ أن مشاهدات عينة سكان الريف تتراوح بين ٩٥ ، ٢٣١ وهو مدى يساوي ٢٣١ ـ ٩٥ = ١٣٦ ، بينما تتراوح مشاهدات عينة سكان الحضر بين ١٣٣ ، ٢٨٤ وهو مدى يساوي ٢٨٤ ـ ١٣٣ = ١٥١ . ويلاحظ كذلك عدم وجود اختلافات أساسية في شكل التوزيعين فكلاهما ذو قمة واحدة قريبة

من مركز التوزيع . إنما يختلف التوزيعان أساساً في مكان مركز كل منهما حيث يقع مركز التوزيع لعينة سكان الريف عند الغصن ١٤ بينما يأتي مركز توزيع عينة سكان الحضر عند الغصن ٢٠ . أي أن مشاهدات توزيع سكان الحضر تزيد في المتوسط بمقدار ٦٠ وحدة تقريباً عن مشاهدات توزيع سكان الريف. ويمكن تفسير هذه الاختلافات بما هو معروف من ارتفاع نسبة الدهون في غذاء الموسرين من سكان المدن . وسنرى فيما بعد كيف يمكن حساب مقاييس عددية لاستنتاج الفروق بين مركزي مثل هذين التوزيعين بشكل أكثر دقة .

٦ ـ التوزيع التكراري للمتغيرات الكمية المتصلة

تختلف عملية إنشاء التوزيع التكراري في حالة المتغيرات المتصلة عن غيرها في أمرين ، الأول هو عدم وجود قيم أو فئات طبيعية للمتغيرات المتصلة يمكن أن تستخدم كأساس لتصنيف البيانات على عكس الوضع في حالة البيانات النوعية والبيانات المتقطعة ، ولا بد تبعاً لذلك من وضع المعايير اللازمة لإنشاء هذه الفئات. أما الأمر الثاني ،فهو أن البيانات المتصلة كما هو معروف تنتج من عملية قياس وأن المشاهدات تكون تبعاً لذلك قيماً تقريبية . وتعتمد درجة التقريب في هذه المشاهدات على مدى كفاءة ودقة جهاز القياس المستخدم من ناحية وعلى درجةالدقة المطلوبة في النتائج من ناحية أخرى . ويجب أن يؤخذ ذلك في الاعتبار عند تفسير المشاهدات، فمثلاً اذا كان طول الشخص مقاساً لأقرب سنتمتر فإن المشاهدة ١٧٠ سم تعني أن الطول الفعلي يقع بين ١٦٩,٥ ، ١٧٠,٥ سم أما إذا كان الطول مقاساً لأقرب جزء عشري من السنتمتر فإن المشاهدة ١٧٠,٣ سم تعني أن الطول الفعلي يقع بين ١٧٠,٢٥ ، ١٧٠,٣٥ سم ، كذلك إذا كان الطول مقاساً لأقرب جزء مئوي من السنتمتر فإن المشاهدة ١٧٠,٣٤ سم تعني أن الطول الفعلي يقع بين ١٧٠,٣٣٥ ، ١٧٠,٣٤٥ سم ، وهكذا .

ولعل أكثر الأمور أهمية عند انشاء التوزيع التكراري للمتغير المتصل هو

تحديد فئات المتغير التي تتخذ أساساً لهذا التوزيع . ويتطلب ذلك تحديد عدد هذه الفئات وتحديد المدى الذي ستشمله كل فئة . ويعطي جدول (١٤) مثالاً لتوزيع تكراري لمتغير متصل ، اذ يظهر التوزيع العمري لمفردات عينة حجمها ١٠٠ من موظفي الدولة .

جدول (١٤)

التوزيع العمري في عينة من مائة موظف

فئات العمر بالسنوات	عدد الموظفين
أقل من ٢٠	٢
٢٠ ـ ٢٩	٢٠
٣٠ ـ ٣٩	٣٥
٤٠ ـ ٤٩	٢٥
٥٠ ـ ٥٩	١٥
٦٠ فأكثر	٣
المجموع	١٠٠

(المصدر : بيانات افتراضية)

تجدر الإشارة إلى أن اختيار الفئات التي تستخدم في التوزيع أمر تحكمي إلى حد ما ، ولكنه يعتمد بدرجة كبيرة على طبيعة المشاهدات وعلى الهدف الذي من أجله يتم انشاء التوزيع . وهناك بعض القواعد العامة التي يمكن الاسترشاد بها في هذا الصدد بهدف التقليل من تأثير الطبيعة التحكمية لهذا الاختيار .

١ ـ يجب أن يكون عدد الفئات كافياً لتوضيح معالم التوزيع بشكل جيد . وفي هذا الصدد ، ينصح ألا يقل عدد هذه الفئات عن ٦ وألا يزيد عن

١٥ فئة . ويعتمد العدد المختار على حجم البيانات ، اذ يمكن زيادة عدد الفئات المستخدمة كلما كان عدد المشاهدات كبيراً .

٢ ـ يجب التأكد من أن كل مشاهدة سوف تصنف في فئة واحدة فقط من هذه الفئات . ويتطلب ذلك ضرورة أن تشمل الفئات المختارة على أصغر مشاهدة وأكبر مشاهدة ، والتأكد من عدم وجود فجوات بين الفئات المتتالية بالإضافة إلى عدم حدوث تداخل بين هذه الفئات (أي عدم وجود قيم مشتركة بينها) . ويجب ، في هذا الصدد ، تفسير حدود الفئات في ضوء نظام التقريب المتبع في تسجيل المشاهدات . فمثلاً يقاس العمر في جدول (١٤) بالسنوات مما يعني أن الحدود الفعلية للفئة الأولى هي ١٩,٥ فأقل وللفئة الثانية ١٩,٥ ـ ٢٩,٥ وللفئة الثالثة ٢٩,٥ ـ ٣٩,٥ وهكذا .

٣ ـ يفضل أن تكون أطوال الفئات متساوية بقدر الإمكان . ويحسن أن تأخذ هذه الأطوال قيماً يسهل التعامل معها مثل ٥ ، ١٠ ، ٥٠ ، ١٠٠ ، . . . الخ . لأن ذلك يسهل عمليات انشاء وقراءة واستخدام التوزيع التكراري .

٤ ـ يلاحظ أن الفئة الأولى في جدول (١٤) هي فئة مفتوحة من أسفل ، بينما الفئة الأخيرة في نفس الجدول فئة مفتوحة من أعلى . وتستخدم هذه الفئات المفتوحة عندما يكون عدد المفردات في هذه الفئات قليلاً ويراد تقليل عدد الفئات المستخدمة لتغطيتها . ويجب كقاعدة عامة تفادي استخدام هذه الفئات المفتوحة كلما كان ذلك ممكناً . وسنرى فيما بعد أن استخدام هذه الفئات يترتب عليه صعوبات في حساب بعض المقاييس الإحصائية .

ويوضح المثال التالي كيفية التطبيق العملي لهذه القواعد . يعطي جدول (١٥) عدد الساعات التي يقضيها الطالب أسبوعياً في ممارسة هواياته لمفردات عينة حجمها ٨٠ من طلبة الجامعة .

جدول (١٥)

عدد الساعات التي يقتضيها الطالب اسبوعياً في ممارسة هواياته لعينة من ٨٠ طالب

٢٣	٢٤	١٨	١٤	٢٠	٢٤	٢٤	٢٦	٢٣	٢١
١٦	١٥	١٩	٢٠	٢٢	١٤	١٣	٢٠	١٩	٢٧
٢٩	٢٢	٣٨	٢٨	٣٤	٣٢	٢٣	١٩	٢١	٣١
١٦	٢٨	١٩	١٨	١٢	٢٧	١٥	٢١	٢٥	١٦
٣٠	١٧	٢٢	٢٩	٢٩	١٨	٢٥	٢٠	١٦	١١
١٧	١٢	١٥	٢٤	٢٥	٢١	٢٢	١٧	١٨	١٥
٢١	٢٠	٢٣	١٨	١٧	١٥	١٦	٢٦	٢٣	٢٢
١١	١٦	١٨	٢٠	٢٣	١٩	١٧	١٥	٢٠	١٠

يتطلب وضع هذه البيانات في جدول تكراري البدء بتحديد اصغر قيمة واكبر قيمة في البيانات . هذه القيم هي ١٠ ، ٣٨ على الترتيب وهو ما يعني أن المدى = ٣٨ - ١٠ = ٢٨ . يتم بعد ذلك تقرير عدد الفئات التي تستخدم في التوزيع . اذا اتفق على أن يكون عدد الفئات يساوي ٦ مثلاً فإن معنى ذلك أن طول كل فئة هو ٢٨/٦ = ٥ تقريباً . وتكون الفئات في هذه الحالة هي ١٠ - ١٤ ، ١٥ - ١٩ ، ٢٠ - ٢٤ ، ٢٥ - ٢٩ ، ٣٠ - ٣٤ ، ٣٥ - ٣٩ .

ويلاحظ أن هذه الفئات متساوية الطول ، حيث طول كل منها = ٥ . كذلك تبدأ الفئات وتنتهي عند أرقام يسهل التعامل معها . هذا بالإضافة إلى عدم وجود فجوات أو تداخل بين الفئات المتتالية . ولما كانت البيانات مقاسة لأقرب ساعة ، فإن الحدود الفعلية لهذه الفئات هي ٩,٥ - ١٤,٥ ، ١٤,٥ - ١٩,٥ ، ١٩,٥ - ٢٤,٥ ، ٢٤,٥ - ٢٩,٥ ، ٢٩,٥ - ٣٤,٥ ، ٣٤,٥ - ٣٩,٥ ، على الترتيب . وينشأ التوزيع التكراري المطلوب بتصنيف المشاهدات على هذه الفئات . ويظهر جدول (١٦) نتيجة هذا التصنيف ، مع اعطاء التكرارات النسبية المناظرة .

جدول (١٦)

التوزيع التكراري لمفردات عينة من ٨٠ طالباً حسب عدد الساعات التي يقضيها كل منهم في ممارسة هواياته

فئات الزمن بالساعات	الحدود الفعلية للفئات	عدد الطلبة (التكرارات)	التكرار النسبي
١٠ - ١٤	٩,٥ - ١٤,٥	٨	٨/٨٠ = ٠,١٠
١٥ - ١٩	١٤,٥ - ١٩,٥	٢٨	٢٨/٨٠ = ٠,٣٥
٢٠ - ٢٤	١٩,٥ - ٢٤,٥	٢٧	٢٧/٨٠ = ,٣٤
٢٥ - ٢٩	٢٤,٥ - ٢٩,٥	١٢	١٢/٨٠ = ,١٥
٣٠ - ٣٤	٢٩,٥ - ٣٤,٥	٤	٤/٨٠ = ,٠٥
٣٥ - ٣٩	٣٤,٥ - ٣٩,٥	١	١/٨٠ = ٠,٠١
المجموع		٨٠	١,٠٠

(المصدر : جدول (١٥))

يوضح التوزيع التكراري نمط الاختلاف في البيانات . وقد سبقت الاشارة إلى أن درجة التفصيل المتاحة في هذه التوزيعات أقل من درجة التفصيل في البيانات الأصلية ، ذلك أن التوزيع التكراري يهمل التفاصيل الخاصة بالمشاهدات داخل كل فئة . فمثلاً ، يتضح من جدول (١٦) أن هناك ثمانية طلبة يقضون ما بين ١٠ ساعات ، ١٤ ساعة أسبوعياً في ممارسة هواياتهم وذلك دون إعطاء أية تفاصيل عن العدد الفعلي للساعات التي يقضيها كل منهم .

ويمكن للباحث استخدام شكل الأغصان والأوراق لعرض بياناته اذا دعت الحاجة إلى دراسة التفاصيل الخاصة بالمشاهدات التي تقع داخل كل فئة . وفي هذا الصدد ، يجب ملاحظة العلاقة الوثيقة بين التوزيع التكراري

وبين شكـل الأغصان والأوراق لنفس البيانـات ، اذ يمكن النـظر إلى فئـات الجدول على أنها أغصان وإلى التكرارات على أنها أعداد الأوراق على هـذه الأغصان .

يتبقى بعـد ذلك ضـرورة الإشارة إلى بعض المـواقف التي قد تتـطلب استخدام فئات غير متساوية الطول كأساس لانشاء التوزيع التكراري . ويوضح شكل (٦) مثالاً لأحد هذه المواقف ، حيث تتركز الغالبية العظمى للمشاهدات داخل مدى ضيق ، بينما يوجد عدد قليل من المشاهدات خارج هذا المدى .

إذا تقـرر استخدام فئـات ضيقة متسـاوية الـطول ، فـإن عـدد الفئـات المطلوبة سيكون كبيراً جداً بالإضـافة إلى أن الكثيـر من هذه الفئـات سيكون خالياً من المشاهدات كما يظهر في شكل ٦ (أ) . كـذلك اذا تقـرر استخدام فئات واسعة متساوية الطول فإن معظم المشاهدات سوف تتركز في عدد قليل من هذه الفئات بحيث يصبح التعرف على نمط الاختـلاف في البيانـات أمراً صعباً ، كما يبـدو في شكل ٦ (ب) . قـد يفضـل في مثـل هـذه الحـالات استخدام فئات ضيقـة متساويـة الطول داخـل المدى الـذي تتركـز فيه معـظم المشاهدات واستخدام فئات واسعة خارج هذا المدى ، كما يوضـح شكل ٦ (حـ) . ويعـطي جدول (١٧) مثـالاً لتوزيـع تكراري ذو فئـات غير متسـاوية الطول .

(أ)

(ب)

(جـ)

شكل (٦) : طرق مختلفة لاختيار الفئات في حالة تركـز معظم المشـاهدات داخـل مدى ضيق .

(أ) عدد كبير من الفئات الضيقة متساوية الطول .

(ب) عدد قليل من الفئات الواسعة متساوية الطول .

(جـ) فئات غير متساوية .

جدول (١٧)

توزيع الأجر الشهري لعينة من ١٠٠ من موظفي الدولة

فئات الأجر بالدراهم	عدد الموظفين
أقل من ٣٠٠٠	٤
٣٠٠٠ ـ ٣٩٩٩	٢٠
٤٠٠٠ ـ ٤٩٩٩	٣٠
٥٠٠٠ ـ ٥٩٩٩	٢٥
٦٠٠٠ ـ ٧٩٩٩	١٥
٨٠٠٠ فأكثر	٦
المجموع	١٠٠

المصدر : (بيانات افتراضية) .

وتجدر الاشارة إلى أن الحاسبات الآلية تستخدم بشكل واسع لتصنيف وعرض البيانات الاحصائية ، خاصة اذا كان حجم مجموعة البيانات كبيراً . وقد سبقت الاشارة إلى بعض مجموعات البرامج المتاحة لاجراء العمليات الاحصائية المختلفة . ويقوم الباحثون عند استخدام هذه البرامج بإدخال بياناتهم بالطريقة المناسبة ، ثم يطلب إلى الحاسب الآلي تنفيذ عمليات معينة وفق تعليمات محددة . ويمكن استخدام الحاسب الآلي من تجربة أساليب مختلفة لتصنيف البيانات ، والنظر إلى أنماط متعددة للفئات ، يستقر بعدها الباحث على شكل نهائي ملائم لتحقيق الأهداف المرجوة .

٧ ـ التوزيع التكراري التجميعي

هناك أسلوبان أساسيان لاعادة عرض المعلومات المتضمنة في جدول التوزيع التكراري . الأسلوب الأول ، هو انشاء التوزيع التكراري النسبي ، وقد سبقت الاشارة الى فوائد هذا التوزيع في التعرف على الأهمية النسبية للأوجه والفئات المختلفة للمتغير وفي اجراء المقارنات بين التوزيعات

المختلفة . أما الأسلوب الثاني فيتمثل في تكوين التوزيع التكراري التجميعي المناظر للبيانات . ويستخدم هذا التوزيع كأداة للوصف والتحليل في المواقف التي تتضمن ترتيب المشاهدات تصاعدياً أو تنازلياً . فمثلاً عند دراسة الدخل الشهري للأسرة ، قد يراد التعرف على عدد الأسر التي تحصل على دخل يقل عن عشرة آلاف درهم ، كذلك عند تحليل انماط الدرجات التي حصل عليها طلبة الثانوية العامة قد يكون من المفيد تحديد عدد الطلبة الذين تزيد درجاتهم عن ٩٠ درجة ، وعند دراسة أعمار الأفراد قد يراد التعرف على نسبة السكان الذين يعيشون حتى يبلغ عمرهم ٧٠ سنة ، وهكذا .

ويمكن التمييز بين نوعين من التوزيعات التجميعية . النوع الأول هو التوزيع التكراري التجميعي الصاعد حيث تحسب التكرارات بتجميع

جدول (١٨)

التوزيع التجميعي الصاعد لعينة من ٨٠ طالباً حسب عدد الساعات التي يقضونها في ممارسة هواياتهم

عدد الساعات	التكرار التجميعي الصاعد	التكرار التجميعي النسبي
أقل من ١٠	صفر	صفر/٨٠ = صفر
أقل من ١٥	صفر + ٨ = ٨	٨/٨٠ = ٠,١٠
أقل من ٢٠	٨ + ٢٨ = ٣٦	٣٦/٨٠ = ٠,٤٥
أقل من ٢٥	٣٦ + ٢٧ = ٦٣	٦٣/٨٠ = ٠,٧٩
أقل من ٣٠	٦٣ + ١٢ = ٧٥	٧٥/٨٠ = ٠,٩٤
أقل من ٣٥	٧٥ + ٤ = ٧٩	٧٩/٨٠ = ٠,٩٩
أقل من ٤٠	٧٩ + ١ = ٨٠	٨٠/٨٠ = ١,٠٠

(المصدر : جدول (١٦)) .

جدول (١٩)

التوزيع التجميعي الهابط لعينة من ٨٠ طالباً حسب عدد الساعات التي يقضونها في ممارسة هواياتهم

عدد الساعات	التكرار التجميعي الهابط	التكرار التجميعي النسبي
١٠ أو أكثر	٨ + ٧٢ = ٨٠	$\frac{٨٠}{٨٠}$ = ١,٠٠
١٥ أو أكثر	٢٨ + ٤٤ = ٧٢	$\frac{٧٢}{٨٠}$ = ٠,٩٠
٢٠ أو أكثر	٢٧ + ١٧ = ٤٤	$\frac{٤٤}{٨٠}$ = ٠,٥٥
٢٥ أو أكثر	١٢ + ٥ = ١٧	$\frac{١٧}{٨٠}$ = ٠,٢١
٣٠ أو أكثر	٤ + ١ = ٥	$\frac{٥}{٨٠}$ = ٠,٠٦
٣٥ أو أكثر	١ + صفر = ١	$\frac{١}{٨٠}$ = ٠,٠١
٤٠ أو أكثر	صفر	$\frac{صفر}{٨٠}$ = ٠,٠٠

التكرارات المتتالية بدءاً من أول تكرار في الجدول الأصلي ، والنوع الثاني هو التوزيع التكراري التجميعي الهابط حيث تحسب التكرارات فيه بتجميع التكرارات المتتالية بدءاً من آخر تكرار في الجدول الأصلي .

ويمكن أيضاً تكوين التوزيع التكراري التجميعي النسبي سواء بقسمة التكرارات التجميعية على العدد الكلي للمشاهدات أو بتجميع التكرارات النسبية في الجدول الأصلي مباشرة .

يعطي جدول (١٨) التوزيع التكراري التجميعي الصاعد المناظر لجدول (١٦) ويعطي جدول (١٩) التوزيع التكراري التجميعي الهابط المناظر لنفس البيانات . ويلاحظ أن قيم العمود الأول في الجدول التجميعي الصاعد تنشأ بكتابة « أقل من » أمام جميع « الحدود الدنيا » للفئات في الجدول

الأصلي بينما تكتب كلمة « أو أكثر » بعد جميع هذه الأرقام للحصول على قيم العمود الأول في الجدول التجميعي الهابط . ويلاحظ كذلك ان تعريف التكرارات الصاعدة والتكرارات الهابطة يؤدي الى ان يتساوى مجموع هذين التكرارين المناظرين لقيمة محددة لعدد الساعات مع العدد الكلي للمشاهدات . فمثلاً عند القيمة ١٠ ساعات نجد أن هذا المجموع هو صفر + ٨٠ = ٨٠ وعند القيمة ١٥ ساعة يكون المجموع ٨+٧٢=٨٠ ، وهكذا .

٨ ـ التوزيع التكراري المشترك لعدد من المتغيرات

يلجأ الباحثون عند محاولة شرح وتفسير نمط الاختلاف المشاهد في متغير ما الى دراسة العوامل المرتبطة بهذا المتغير . مثال ذلك دراسة العلاقة بين عمر الزوج وعمر الزوجة ، أو بين كمية الطلب على سلعة ما وكمية المعروض منها ، أو بين درجة انتشار الجريمة في المناطق المختلفة ومستوى دخول الأسر في هذه المناطق ، أو بين نوع الطالب والدرجة التي يحصل عليها في مبادىء الاحصاء ، . . . الخ . ويتم في هذه الحالات جمع بيانات عن المتغيرات ، ثم استخدام الأساليب الاحصائية لتحليل العلاقات المشاهدة في هذه البيانات .

ويعتبر وضع البيانات في شكل توزيع تكراري مشترك الخطوة الأولى في وصف ودراسة طبيعة العلاقة بين المتغيرات الاحصائية . ويتم انشاء هذا التوزيع باتباع الخطوات المعتادة حيث تتحدد أولاً الأوجه أو الفئات المختلفة لكل متغير ، ثم تصنف المشاهدات بعد ذلك على أوجه أو فئات هذه المتغيرات في آن واحد .

يعطي جدول (٢٠) مثالاً لمجموعة بيانات عن متغيرين هما نوع الشخص (ذكر ، أنثى) وحيازة الشخص لرخصة قيادة (نعم ، لا) ، لمفردات عينة من ٢٥ شخصاً من سكان المدينة .

جدول (٢٠)

بيانات عن نوع الشخص وحيازته لرخصة قيادة ، لمفردات عينة من ٢٥ شخصاً

رقم الشخص	النوع	حيازة الرخصة	رقم الشخص	النوع	حيازة الرخصة
١	ذكر	نعم	١٤	انثى	لا
٢	ذكر	نعم	١٥	ذكر	نعم
٣	انثى	لا	١٦	ذكر	نعم
٤	ذكر	لا	١٧	ذكر	لا
٥	ذكر	نعم	١٨	ذكر	نعم
٦	ذكر	نعم	١٩	ذكر	نعم
٧	ذكر	نعم	٢٠	انثى	نعم
٨	ذكر	لا	٢١	انثى	نعم
٩	ذكر	نعم	٢٢	انثى	لا
١٠	انثى	لا	٢٣	انثى	لا
١١	انثى	لا	٢٤	انثى	نعم
١٢	انثى	لا	٢٥	انثى	لا
١٣	انثى	نعم			

المصدر : بيانات افتراضية .

ويظهر التوزيع التكراري للمتغيرين في آن واحد لمفردات العينة في جدول (٢١) ، حيث يلاحظ أن هناك عشرة ذكور لديهم رخصة قيادة وثلاثة ذكور بدون هذه الرخصة وأن هناك أربع اناث لدى كل منهم رخصة قيادة وثمان اناث بدون هذه الرخصة . يلاحظ كذلك ان مفردات العينة موزعة حسب النوع الى ١٣ ذكراً ، ١٢ أنثى ، كما أن ١٤ شخصاً في العينة لديهم رخصة قيادة ، ١١ شخصاً ليس لديهم هذه الرخصة .

جدول (٢١)

التوزيع التكراري المشترك للنوع وحيازة رخصة القيادة في عينة من ٢٥ شخصاً

رخصة القيادة / النوع	نعم	لا	المجموع
ذكر	١٠	٣	١٣
أنثى	٤	٨	١٢
المجموع	١٤	١١	٢٥

المصدر : جدول (٢٠)

يعطي جدول (٢٢) مثالاً آخر لمجموعة بيانات مزدوجة ، حيث تظهر كمية الدجاج وكمية لحوم البقر والأغنام (كلاهما مقاس بالكيلوجرام) المستهلكة في أحد المطاعم الكبيرة خلال ٢٤ يوماً متتالياً :

جدول (٢٢)

كمية الدجاج وكمية اللحوم المستهلكة في أحد المطاعم خلال ٢٤ يوماً

كمية الدجاج	كمية اللحوم	كمية الدجاج	كمية اللحوم
٦٥	٤١	٦٤	٣٥
٧١	٣٧	٦٥	٤٠
٧٤	٤٨	٧٨	٥١
٧٥	٤٦	٧٦	٤٩
٧٢	٤٥	٦٧	٤٣
٧٨	٤٣	٧٢	٤٢

كمية الدجاج	كمية اللحوم	كمية الدجاج	كمية اللحوم
٧٣	٥٣	٧٦	٤١
٧٥	٤٦	٦٠	٤٣
٦٧	٣٦	٧١	٤٥
٦٣	٣٨	٦٦	٤٨
٧٠	٤٨	٧٦	٤٦
٧٤	٤٧	٧٢	٤١

المصدر : بيانات فرضية .

يلاحظ عند إنشاء جدول التوزيع التكراري المشترك المناظر لهذه البيانات أن ٦٠ هو الحد الأدنى لكمية الدجاج المستهلكة يومياً وأن ٧٨ هو الحد الأعلى . كذلك فإن ٣٥ هو الحد الأدنى لكمية اللحوم المستهلكة يومياً وأن ٥٣ هو الحد الأعلى . إذا تقرر تبعاً لذلك استخدام الفئات ٦٠ ـ ٦٤ ، ٦٥ ـ ٦٩ ، ٧٠ ـ ٧٤ ، ٧٥ ـ ٧٩ كفئات لكمية الدجاج المستهلكة ، واستخدام الفئات ٣٥ ـ ٣٩ ، ٤٠ ـ ٤٤ ، ٤٥ ـ ٤٩ ، ٥٠ ـ ٥٤ كفئات لكمية اللحوم المستهلكة ، ثم صنفت المشاهدات تبعاً لهذه الفئات فإن التوزيع المشترك الناتج يظهر في جدول (٢٣) ، حيث يتضح عدد الأيام التي تقع في كل خلية من خلايا الجدول . فمثلاً هناك يومان يستهلك فيهما كمية من الدجاج بين ٦٠ ، ٦٤ كجم وكمية من اللحوم بين ٣٥ ، ٣٩ كجم .

وفيما يلي بعض الملاحظات الهامة التي يجب ان تؤخذ في الاعتبار عند دراسة جداول التوزيعات التكرارية المشتركة .

أ ـ الهوامش أو التوزيعات الهامشية : يمكن استنتاج التوزيع التكراري لكل متغير على حدة من جدول التوزيع التكراري المشترك وذلك باستخدام

جدول (٢٣)

التوزيع المشترك لكمية الدجاج وكمية اللحوم المستهلكة يومياً في المطعم خلال ٢٤ يوماً متتالياً

كمية الدجاج / كمية اللحوم	٦٠ ـ ٦٤	٦٥ ـ ٦٩	٧٠ ـ ٧٤	٧٥ ـ ٧٩	المجموع
٣٥ ـ ٣٩	٢	١	١		٤
٤٠ ـ ٤٤	١	٣	٢	٢	٨
٤٥ ـ ٤٩		١	٥	٤	١٠
٥٠ ـ ٥٤			١	١	٢
المجموع	٣	٥	٩	٧	٢٤

(المصدر : جدول ٢٢)

هوامش هذا الجدول . فمثلاً يلاحظ ان أول عمود وآخر عمود في جدول (٢١) يعطيان التوزيع التكراري لمفردات العينة حسب النوع فقط كما ان أول سطر وآخر سطر يمثلان التوزيع التكراري حسب حيازة رخصة القيادة فقط . وبنفس الطريقة ، فان أول عمود وآخر عمود في جدول (٢٣) يوضحان التوزيع التكراري لكمية اللحوم المستهلكة يومياً بينما يمثل أول سطر وآخر سطر في هذا الجدول التوزيع التكراري لكمية الدجاج المستهلكة يومياً . ويطلق عادة على هذه التوزيعات اسم التوزيعات الهامشية Marginal distributions.

ب ـ **حساب التوزيعات النسبية** . يمكن حساب العديد من التوزيعات التكرارية النسبية من جدول التوزيع التكراري المشترك . إذ يمكن مثلاً حساب التوزيع النسبي داخل كل صف من صفوف الجدول أو داخل كل عمود من أعمدته ، كما يمكن حساب التوزيع التكراري النسبي المشترك في الجدول ككل . ويتم اختيار بعض أو كل هذه التوزيعات في ضوء طريقة جمع البيانات

والأهداف المحددة للدراسة . فمثلاً ، اذا كانت البيانات في جدول (٢١) قد جمعت بهدف الإجابة عن السؤال : « هل يختلف الذكور عن الإناث في مدى حيازتهم لرخصة قيادة ؟ » فإن الأسلوب المناسب في هذه الحالة هو المقارنة بين التوزيعات النسبية داخل كل صف من صفوف الجدول ، كما يظهر في جدول (٢٤) .

جدول (٢٤)

التوزيع النسبي للذكور والإناث حسب حيازة رخصة القيادة

رخصة القيادة / النوع	نعم	لا	المجموع
ذكر	٧٧٪	٢٣٪	١٠٠٪
أنثى	٣٣٪	٦٧٪	١٠٠٪
جميع الأشخاص	٥٦٪	٤٤٪	١٠٠٪

(المصدر : جدول (٢١))

ويلاحظ في هذا الجدول ان غالبية الذكور (٧٧٪) لديهم رخصة قيادة ، على حين أن أقلية من الإناث لديهم هذه الرخصة . ويمكن القول ان الجدول يوضح أن الذكر أكثر احتمالاً لحيازة رخصة قيادة من الأنثى ، أي أن هناك علاقة بين نوع الشخص ومدى حيازته لرخصة قيادة . ويقال في مثل هذه الحالات أن هناك ارتباطاً Association بين المتغيرين .

أما إذا كانت البيانات في جدول (٢١) قد جمعت بهدف الإجابة عن السؤال « هل يختلف التوزيع النوعي للأشخاص الحائزين على رخصة قيادة عن التوزيع النوعي لغيرهم ؟ » فإن الأسلوب المناسب في هذه الحالة هو

حساب التوزيع النوعي النسبي داخل كل عمود من أعمدة الجدول ، كما يظهر في جدول (٢٥) . ويلاحظ في هذا الجدول أن معظم الحائزين على رخصة قيادة ذكوراً (٧١٪) على حين أن معظم غير الحائزين على رخصة إناثاً (٧٣٪) . أي أن الشخص الحائز على رخصة قيادة اكثر احتمالاً في أن يكون ذكراً مما يعني ان هناك ارتباطاً بين المتغيرين ، وتكون الإجابة عن السؤال المطروح هي بالايجاب .

جدول (٢٥)

التوزيع النوعي للأشخاص الحائزين والأشخاص غير الحائزين على رخصة قيادة

رخصة القيادة / النوع	نعم	لا	جميع الأشخاص
ذكر	٧١٪	٢٧٪	٥٢٪
أنثى	٢٩٪	٧٣٪	٤٨٪
المجموع	١٠٠٪	١٠٠٪	١٠٠٪

(المصدر : جدول (٢١))

وإذا كانت البيانات قد جمعت بهدف الإجابة عن السؤال : « ما هو شكل التوزيع التكراري المشترك للنوع وحيازة رخصة القيادة ؟ » فإن الأسلوب المناسب في هذه الحالة هو حساب نسبة التكرار في كل خلية من خلايا الجدول الى العدد الكلي للمشاهدات . ويظهر ذلك في جدول (٢٦) .

جدول (٢٦)
التوزيع التكراري النسبي المشترك لمفردات العينة حسب النوع وحيازة رخصة القيادة

رخصة القيادة / النوع	نعم	لا	المجموع
ذكر	٤٠٪	١٢٪	٥٢٪
أنثى	١٦٪	٣٢٪	٤٨٪
المجموع	٥٦٪	٤٤٪	١٠٠٪

(المصدر : جدول (٢١))

ويلاحظ أن ٤٠٪ من مفردات العينة هم من الذكور الحائزين على رخصة قيادة وأن ١٢٪ من هذه المفردات هم من الذكور غير الحائزين على رخصة قيادة وأن ١٦٪ من مفردات العينة هم اناث حائزين على رخصة قيادة بينما أن ٣٢٪ من هذه المفردات هم اناث غير حائزات على رخصة قيادة . كذلك فان نسبة الذكور في العينة تبلغ ٥٢٪ وأن نسبة الاناث تبلغ ٤٨٪ .

يجب عند استخدام النسب لوصف التوزيع التكراري أن تعطى البيانات الأصلية التي حسبت منها هذه النسب حتى يتمكن القارىء من تفسيرها على نحو صحيح . ويمكن تحقيق ذلك بوضع النسب في أقواس بالصورة الموضحة في جدول (٢٧) . يعطي جدول (٢٧) توزيع معدلات الجريمة في ٨٠ حياً سكنياً حسب مستوى متوسط الدخل في هذه الأحياء .

جدول (٢٧) توزيع معدلات الجرائم في ٨٠ حياً سكنياً حسب متوسط الدخل في هذه الأحياء

متوسط الدخل / معدل الجريمة	منخفض	متوسط	مرتفع	جميع الدخول
منخفض	١ (١٠٪)	٤ (١٠٪)	٣ (١٠٪)	٨ (١٠٪)
متوسط	٧ (٧٠٪)	٢٨ (٧٠٪)	٢١ (٧٠٪)	٥٦ (٧٠٪)
مرتفع	٢ (٢٠٪)	٨ (٢٠٪)	٦ (٢٠٪)	١٦ (٢٠٪)
المجموع	١٠ (١٠٠٪)	٤٠ (١٠٠٪)	٣٠ (١٠٠٪)	٨٠ (١٠٠٪)

(المصدر : بيانات افتراضية)

ويلاحظ في هذا الجدول ان توزيع معدل الجريمة ثابت لمستويات الدخل المختلفة ، أي أن اختلاف مستوى الدخل في الأحياء المختلفة لا يؤثر على نمط معدل الجريمة في هذه الأحياء . ويقال في هذه الحالة ان معدل الجريمة في الحي مستقل عن متوسط الدخل في الحي .

إذا حسبت التوزيعات النسبية في جدول (٢٧) في الاتجاه الآخر ، أي داخل كل صف من صفوف الجدول فإن النسب الناتجة تظهر في جدول (٢٨) .

جدول (٢٨)

توزيع متوسط الدخل في ٨٠ حياً سكنياً حسب معدل الجريمة في هذه الأحياء

متوسط الدخل / معدل الجريمة	منخفض	متوسط	مرتفع	المجموع
منخفض	١ (١٢٪)	٤ (٥٠٪)	٣ (٣٨٪)	٨ (١٠٠٪)
متوسط	٧ (١٢٪)	٢٨ (٥٠٪)	٢١ (٣٨٪)	٥٦ (١٠٠٪)
مرتفع	٢ (١٢٪)	٨ (٥٠٪)	٦ (٣٨٪)	١٦ (١٠٠٪)
جميع مستويات الجريمة	١٠ (١٢٪)	٤٠ (٥٠٪)	٣٠ (٣٨٪)	٨٠ (١٠٠٪)

(المصدر : جدول (٢٧))

ويوضح جدولي (٢٧) ، (٢٨) حقيقة إحصائية أساسية وهي أنـه اذا كان توزيع المتغير س ثابت لجميع مستويات المتغير ص فإن توزيع المتغير ص يكون أيضاً ثابتاً لجميع مستويات المتغير س ويحدث ذلك دائماً عندما يكون المتغيـرين س ، ص مستقلين كما هـو الحال في هـذا المثال حيث نجـد أن معدل الجريمة ومستوى الدخل مستقلان .

كمثال ثالث ، يعـطي جدول (٢٩) التـوزيع التكـراري المشترك لعمـر الشخص وراتبه الشهري ، مع التوزيع النسبي للرواتب داخل كل فئة عمـرية لمفردات عينة من ٤٠٠ موظف .

جدول (٢٩)

التوزيع التكراري المشترك لعمر الشخص وراتبه الشهري ، والتوزيع النسبي للرواتب داخل كل فئة عمرية لمفردات عينة من ٤٠٠ موظف

العمر بالسنوات / الراتب بالدرهم	٢٠ ـ ٢٩	٣٠ ـ ٣٩	٤٠ ـ ٤٩	٥٠ ـ ٥٩	جميع الأعمار
٦٠٠٠ ـ ٦٩٩٩	٥٠ (٥٠٪)	٢٠ (١٤٪)	١٠ (٨٪)	٤ (١٠٪)	٨٤ (٢١٪)
٧٠٠٠ ـ ٧٩٩٩	٣٠ (٣٠٪)	٧٠ (٥٠٪)	٣٠ (٢٥٪)	٦ (١٥٪)	١٣٦ (٣٤٪)
٨٠٠٠ ـ ٨٩٩٩	٢٠ (٢٠٪)	٥٠ (٣٦٪)	٨٠ (٦٧٪)	٣٠ (٧٥٪)	١٨٠ (٤٥٪)
المجموع	١٠٠ (١٠٠٪)	١٤٠ (١٠٠٪)	١٢٠ (١٠٠٪)	٤٠ (١٠٠٪)	٤٠٠ (١٠٠٪)

(المصدر : بيانات افتراضية)

يلاحظ ان توزيع الرواتب يختلف داخل فئات العمر المختلفة ، ويمكن القول ان الراتب يتجه نحو الارتفاع كلما زاد عمر الموظف .

ويمكن بالطبع حساب توزيعات نسبية اخرى في هذا الجدول ، فيمكن مثلاً قسمة كل تكرار في الجدول على ٤٠٠ لنحصل على التوزيع التكـراري النسبي المشترك والذي يوضح الأهمية النسبية لكل خلية من خلايا الجدول في التوزيع المشترك لأعمار ورواتب الموظفين .

٩ ـ أنواع أخرى من الجداول الإحصائية

هناك أنواع من البيانات يمكن ان تعرض في جداول إحصائية ، دون ان تكون هذه البيانات توزيعات تكرارية . وفيما يلي عرض موجز لهذه الأنواع .

أ ـ السلسلات الزمنية . تعرف السلسلة الزمنية بأنها مجموعة متتالية من القراءات أو المشاهدات التي تؤخذ عادة على فترات زمنية متساوية عن احدى الظواهر . مثال ذلك عدد السكان في الدولة في أول يوليو من كل عام ، عدد المواليد السنوية في الدولة ، عدد التلاميذ الحاصلين على شهادة الثانوية العامة سنوياً ، انتاج النفط السنوي في بلدان الخليج ، . . . الخ . وعلى الرغم من انه يمكن اعتبار السلسلة الزمنية بيانات كمية لأن الزمن متغير كمي إلا أن الخاصية الأساسية لهذه السلسلات الزمنية هو اعتماد بياناتها على الزمن ، ويهدف تحليلها تبعاً لذلك الى دراسة نمط هذا الاعتماد . ويعطي جدول (٣٠) مثالاً لسلسلة زمنية .

جدول (٣٠)

عدد حوادث السيارات التي حدثت على طريق
معين خلال السنوات ١٩٨٠ ـ ١٩٨٥

السنة	١٩٨٠	١٩٨١	١٩٨٢	١٩٨٣	١٩٨٤	١٩٨٥
عدد الحوادث	١٣٠	١٣٢	١٢٠	١٤٠	١٥٥	١٦٠

(المصدر : بيانات افتراضية)

ب ـ البيانات الجغرافية . ويقصد بذلك التوزيع الجغرافي للظواهر المختلفة ، مثل توزيع السكان على مناطق الدولة المختلفة وتوزيع طلبة المدارس على المناطق التعليمية المختلفة وتوزيع عدد السيارات المسجلة على ادارات المرور المختلفة ، . . . الخ . ويمكن النظر الى هذه البيانات

كبيانات نوعية إلا أن التركيز في عرضها يكون على البعد الجغرافي للبيانات . ويعطي جدول (٣١) مثالاً لبيانات جغرافية .

جدول (٣١)

عدد سكان دولة الامارات العربية المتحدة في تعداد ١٩٨٠ لكل إمارة

الامـــــارة	عدد السكان بالآلاف
أبوظبي	٤٥٢
دبي	٢٧٦
الشارقة	١٥٩
عجمان	٣٦
أم القيوين	١٣
رأس الخيمة	٧٤
الفجيرة	٣٢
المجموع	**١٠٤٢**

(المصدر : التعداد العام للسكان ١٩٨٠ الجزء الثالث ـ وزارة التخطيط ، دولة الامـارات ـ ١٩٨٢) .

تمرينات

١ ـ تمثل البيانات الآتية وسيلة الانتقال لزيارة احدى المعالم السياحية لمفردات عينة عشوائية من ٥٠ زائراً :

سيارة   حافلة   طائرة   سيارة   قطار
قطار   قطار   سيارة   حافلة   قطار
طائرة   سيارة   طائرة   سيارة   طائرة
طائرة   سيارة   قطار   طائرة   طائرة
طائرة   سيارة   سيارة   طائرة   سيارة

سيارة   سيارة   حافلة   قطار   طائرة
قطار   سيارة   سيارة   قطار   سيارة
سيارة   طائرة   طائرة   سيارة   سيارة
سيارة   حافلة   سيارة   طائرة   قطار
طائرة   طائرة   سيارة   حافلة   حافلة

أ ـ وضع هذه البيانات في توزيع تكراري مناسب .
ب ـ كون التوزيع التكراري النسبي المناظر .

٢ ـ تعطي البيانات الآتية نتيجة سؤال كل من ٢٥ شخصاً من ملاك السيارات الخاصة عن نوع السيارة التي يقودها كل منهم :

يابانية   أوروبية   أوروبية   يابانية   أمريكية   يابانية
أمريكية   يابانية   أوروبية   يابانية   أمريكية   يابانية
يابانية   أمريكية   أمريكية   أوروبية   أوروبية   يابانية
أوروبية   يابانية   أمريكية   يابانية   أوروبية   أمريكية
أمريكية   أمريكية   يابانية   أوروبية   أمريكية   أمريكية

أ ـ ضع هذه البيانات في شكل توزيع تكراري نسبي .

ب ـ اذا كان من الممكن اعتبار هؤلاء الأشخاص كعينة عشوائية من ملاك السيارات في الدولة عموماً ، فهل يمكنك تقدير نسبة ملاك السيارات في الدولة الذين يقودون سيارات أمريكية ؟

٣ ـ أخذت عينة عشوائية حجمها ١٥٠٠ شخص من السكان في سن العمل في إحدى المدن ، بهدف جمع بيانات عن الحالة العملية لكل منهم . تقرر تكرار الاتصال بمفردات العينة حتى نحصل على البيانات المطلوبة ولرفع معدل استجابتهم للدراسة . يعطي الجدول التالي البيانات التي جمعت بعد زيارة واحدة وتلك التي تم جمعها بعد عشر زيارات لهؤلاء المفردات . والمطلوب التعليق على هذه البيانات .

الحالة العملية	عدد الأشخاص بعد زيارة واحدة	عدد الأشخاص بعد عشرة زيارات
يعمل بانتظام	٧٥	٦٥٠
يعمل بعض الوقت	٢٧	١٦٤
عاطل	١٥	٥٠
متقاعد	٥٤	١٣٨
خارج قوة العمل	٧٩	١٩٨
المجموع	**٢٥٠**	**١٢٠٠**

٤ ـ بلغ عدد محاضرات مبادىء الإحصاء لشعبة ما خلال الفصل الدراسي الماضي ٤٠ محاضرة . فيما يلي عدد الطلبة الذين تغيبوا خلال هذه المحاضرات .

٢	صفر	١	صفر	٣	١	١	صفر	١	٣
١	٣	٤	١	٦	صفر	٣	٢	٢	١
٥	٢	٣	٢	٢	٢	٤	١	٤	٢
صفر	١	١	١	٢	١	١	صفر	١	٣

أ ـ ضع هذه البيانات في شكل توزيع تكراري مناسب .

ب ـ احسب نسبة المحاضرات التي يقل عدد الغائبين فيها عن ٣ طلبة .

حـ ـ احسب نسبة المحاضرات التي يبلغ عدد الغائبين فيها ما بين ٣ ، ٥ طلبة .

٥ ـ تم قياس ضغط دم الشخص لأقرب ميليمتر لعينة من ٥٠٠ شخص . صنفت هذه البيانات تبعاً لرقم الآحاد فيها فنتج التوزيع التكراري الآتي :

رقم الآحاد	صفر	١	٢	٣	٤	٥	٦	٧	٨	٩	المجموع
عدد الأشخاص	٦٠	صفر	٨٠	صفر	١١٢	٤٠	٨٨	صفر	١٢٠	صفر	٥٠٠

هل تدل هذه البيانات على أن أرقام الآحاد تتكرر بشكل متساو أم لا ؟ وضح سبب إجابتك .

٦ ـ في دراسة عن مدى إقبال المستهلكين على نوع معين من قطع الصابون في كل من المدينتين أ ، ب ، أخذت عينة عشوائية من ٢٥ مستهلكاً في المدينة أ وعينة عشوائية من ٣٠ مستهلكاً في المدينة ب . فيما يلي عدد قطع الصابون التي استهلكها أفراد العينة من هذا النوع خلال الشهر الماضي :

المدينة (أ)					المدينة (ب)					
صفر	١	١	٢	٢	٤	٣	٦	٩	٢	١
٢	٨	٢	٩	٣	٥	صفر	٥	٩	٥	٢
٥	صفر	٢	٢	٢	٦	صفر	٧	٨	٤	٦
صفر	٣	٤	صفر	٢	٢	٥	٨	٧	٩	٣
١	١	صفر	١	٢	١	٧	٨	٦	٣	٥

(أ) ضع هذه البيانات في توزيعات تكرارية مناسبة .

(ب) قارن بين نمط استهلاك هذا النوع من الصابون في المدينتين وذلك بافتراض أن كل شخص في المدينتين يستهلك ٩ قطع صابون شهرياً .

٧ ـ فيما يلي التوزيع التكراري لمجموعة من ٩٠٠ شخص يمارسون صيد الأسماك كهواية ، حسب عدد الأسماك التي صادها كل منهم من أحد الأنهار خلال فترة زمنية طولها ساعتان .

عدد الأسماك	صفر	١	٢	٣	٤	٥	٦	٧	٨	٩	المجموع
عدد الصائدين	٥٠٤	٦٥	٦٠	٦٦	٥٣	٥٥	٢٧	٢٥	٢٥	٢٠	٩٠٠

(أ) كون التوزيع التكراري النسبي المناظر .

(ب) هل تدل هذه البيانات على صحة الإنطباع السائد بأن نسبة قليلة من الصائدين يصطادون الجزء الأكبر من الأسماك ، أي أن معظم هواة الصيد لا يفلحون في صيد الكثير من الأسماك ؟ وضح سبب إجابتك .

٨ ـ فيما يلي بيانات عن عدد التلاميذ في ٤٠ مدرسة ابتدائية من مدارس الدولة :

٦٤٤	٥١٢	٤٤٨	٧٣٠	٤٠١
٥٠١	٤٥٨	٧٥٥	٥٦٩	٤١٧
٦٠٣	٥٩٩	٧٩١	٤٠٧	٤٣٣
٤٨٤	٥٥٤	٦٣٤	٥٥٣	٧٢٣
٤٠٥	٤١٩	٦٤٧	٧٩٢	٨٨٥
٤٥٠	٥٠٩	٤٤٠	٤٠٢	٦٢٤
٥٥٩	٧٧٧	٨٥٦	٤٩٢	٤٠٠
٥٦٥	٤٢٤	٤١٧	٥٢٤	٤٦٨

أ ـ رتب هذه البيانات تصاعدياً ، ثم ارسمها على خط أفقي وعلق على الشكل الناتج .

ب ـ كون شكل الأغصان والأوراق المناظر لهذه البيانات باستخدام الأغصان ٤ ، ٥ ، ٦ ، ٧ ، ٨ .

حـ ـ اذا تقرر وضع هذه البيانات في توزيع تكراري ذو خمس فئات تناظر الأغصان المختلفة في الجزء (ب). اكتب هذا التوزيع وبين الحدود الفعلية للفئات المستخدمة .

٩ ـ ما هي البيانات الأصلية المناظرة لكل غصن من الأغصان الآتية :

(أ)	٨، ١، ١، ٥، ٧، ٢، صفر	١
(ب)	٢، صفر، ٣، ٣، ٥	١٢
(حـ)	٠١، ٦٦، ١٨، ٤٥	٣
(د)	٩، ٢، ٢، ٧، صفر	١,٥

١٠ ـ تعطي البيانات التالية معدل الوفاة في ٢٤ حياً سكنياً في دولة ما (المعدلات في الألف)

١٠,١	٨,٢	٩,٩	٨,١	١٠,٩	١٢,٨
١١,٨	٨,٤	٨,٩	٩,٨	٧,٤	٨,٩
٧,٩	٧,٧	١١,٣	١٠,٧	١٠,٨	٦,٦
٩,٧	٤,٩	١٤,٧	٨,٨	٩,١	٩,٤

(أ) رتب هذه البيانات تصاعدياً .

(ب) كون شكلاً للأغصان والأوراق لتمثيل هذه البيانات .

(حـ) استخدم (أ)، (ب) لشرح نمط اختلاف معدلات الوفاة في الأحياء المختلفة .

١١ ـ فيما يلي عدد البيض في ثلاثين من عشش السلاحف التي وجدت على شاطىء معين :

٢٠٦	١٦٧	١٧٥	٢٠٤	١٢٣	١٣٨
١٩٧	١٨٧	١٩٣	١٢٤	١٣٧	١٤١
١٤٢	١٩٢	١٩٧	١٠٩	١٢٦	١٢٧
١٨١	١٧١	١٦٣	١٤٦	١٢٤	١٨٤
١٠١	٢٠١	١٣٣	١٤١	١٥٢	١٣٢

أ ـ ضع هذه البيانات في شكل للأغصان والأوراق .

ب ـ إذا علم أن هذه العشش تنتمي لنوعين من السلاحف ، هل يظهر ذلك في شكل الأغصان والأوراق ؟

١٢ ـ حدد الفئات التي يمكن استخدامها للتوزيع التكراري في كل حالة من الحالات الآتية :

(أ) يتراوح الأجر المدفوع يومياً لمجموعة من عمال الخدمات بين ٢١٥,٢٦ درهم ، ٣١٣,٥٣درهم ، ويراد وضع هذه البيانات في توزيع تكراري ذي ست فئات .

(ب) تتراوح درجة غليان أحد العناصر لأقرب درجة مئوية بين ١٣٦° ، ١٦٨° ، ويراد وضع هذه البيانات في توزيع تكراري ذي سبع فئات .

(حـ) يتراوح مجموع الدرجات التي حصل عليها مجموعة من المتقدمين لشغل إحدى الوظائف بين ١٤٨، ٢٣٦ درجة ويراد وضع البيانات في توزيع تكراري ذي عشر فئات .

١٣ ـ اذكر في كل حالة من الحالات الآتية ما إذا كانت الفئات المستخدمة تحقق الشروط الجيدة الواجب توافرها في التوزيع التكراري :

(أ) استخدمت الفئات : صفر ـ ٥ ، ٦ ـ ١٠ ، ١٢ ـ ١٧ ، ١٨ ـ ٢٣، ٢٣ ـ ٣٠ ، لانشاء التوزيع التكراري لعدد الأيام الممطرة خلال شهر ديسمبر من كل عام .

(ب) استخدمت الفئات : صفر ـ ٣٠، ٣٠ ـ ١٠٠، ١٠٠ أو أكثر ، لانشاء التوزيع التكراري للأجر اليومي للعامل .

(حـ) استخدمت الفئات : ٣ أو أقل ، ٤ أو أكثر ، لإنشاء التوزيع التكراري لعدد الأطفال في الأسرة .

(ء) استخدمت الفئات : صفر ـ ٨٩، ١٠٠ ـ ١٨٩، ٢٠٠ أو أكثر لانشاء التوزيع التكراري لدرجات حرارة مجموعة من الأفران .

١٤ ـ ما هي الحدود الفعلية للفئات المستخدمة في التوزيع التكراري في كل حالة من الحالات التالية ؟

(أ) استخدمت الفئات : صفر ـ ١٤ ، ١٥ـ ٢٩ ، ٣٠ ـ ٤٤، ٤٥ ـ ٥٩ ، لإنشاء التوزيع التكراري لعدد الطلبة الغائبين يومياً عن مدرسة ما .

(ب) استخدمت الفئات : صفر ـ ٢، ٣ ـ ٥، ٦ ـ ٨، ٩ ـ ١١، ١٢ ـ ١٤ ، ١٥ ـ ١٧ ، لانشاء التوزيع التكراري لعدد الحقائب التي تفقد أسبوعياً على طائرات إحدى شركات الطيران .

(حـ) استخدمت الفئات: ٣٠ ـ ٣٤، ٣٥ ـ ٣٩، ٤٠ ـ ٥٩، ٦٠ ـ ٧٩، ٨٠ فأكثر لانشاء التوزيع التكراري لدرجة حرارة مجموعة من الأشياء مقاسة لأقرب درجة مئوية .

(ء) استخدمت الفئات : صفر ـ ١٩,٩، ٢٠,٠ ـ ٣٩,٩ ، ٤٠,٠ ـ ٥٩,٩ ، ٦٠,٠٠ ـ ٧٩,٩ ، ٨٠ ـ ٩٩,٩ لإنشاء التوزيع التكراري لأوزان مجموعة من الكائنات الحية لأقرب جزء عشري من الجرام .

١٥ ـ تعطى البيانات التالية الدرجات التي حصل عليها ٥٠ طالباً في أحد اختبارات الإحصاء :

١٣٦	٩٢	١١٥	١١٨	١٢١
١١٩	١١٥	١٠١	١٢٩	٨٧
١٢٧	١٠٣	١١٠	١٢٦	١١٨
١٤٦	١٢٦	١١٩	١١٩	١٠٥
١٠٦	١٢٥	١١٧	١٠٢	١٤٦
١٣٧	١٣٢	١٢٠	١٠٤	١٢٥
١٠٨	١١٠	١٣٣	١٣٥	١٢٦

٨٢	١٠٤	١٣٧	١٢٠	٩٥
١٣٢	١٢٦	١١٨	١٠٠	١١٣
١٢٩	١٢٤	١١٣	٩٥	١٤٨

أ ـ ضع هذه المشاهدات في توزيع تكراري باستخدام الفئات : ٨٠ ـ ٨٩ ، ٩٠ ـ ٩٩ ، ١٠٠ ـ ١٠٩، ١١٠ ـ ١١٩، ١٢٠ ـ ١٢٩، ١٣٠ ـ ١٣٩، ١٤٠ ـ ١٤٩ .

ب ـ اوجد التوزيع التكراري النسبي المناظر .

جـ ـ اوجد التوزيع التجميعي الصاعد المناظر .

د ـ أوجد التوزيع التجميعي الهابط المناظر .

١٦ ـ في إحدى الدراسات الطبية ، تمت ملاحظة ٨٩ مريضاً يستخدم كل منهم جهازاً لتنظيم دقات القلب وسجل الزمن الذي ينقضي بين تاريخ وضع الجهاز داخل جسم المريض وتاريخ أول عطل فني يصيبه بالشهور . وتظهر البيانات فيما يلي :

٢٤	٢٠	١٦	٣٢	١٤	٢٢	٢
١٦	١٢	٢٤	١٤	٢٠	١٨	١٤
٢٨	١٨	١٤	١٠	١٢	٢٤	٦
٢٢	٢٤	٢٦	١٨	٢	١٨	١٢
٢٢	٢٤	٢٢	٢٠	٢٤	٢٨	٢٠
١٨	٢٤	١٨	١٦	٦	١٦	١٠
٣٠	٣٤	٢٦	٢٤	٢٢	٢٨	٣٠
١٢	٢٤	٦	١٠	٢٠	٢٦	
١٦	١٨	٢٠	٢٢	٢٤	٢٠	
١٢	١٨	١٦	٣٤	١٨	١٦	
١٢	٨	٢٤	١٠	١٤	١٦	

٢٢	٢٦	٢٠	٦	١٤	٢٤
١٤	١٨	٢٤	٢٢	٢٨	
٢٢	٢٤	٢٢	٣٢	٨	

أ ـ ضع هذه المشاهدات في توزيع تكراري مستخدماً الفئات : صفر ـ ٥ ، ٦ ـ ١١ ، . . .

ب ـ اوجد التوزيع التكراري النسبي المناظر .

جـ ـ اوجد التوزيع التجميعي النسبي الصاعد المناظر ، واستخدمه في إيجاد :

(١) نسبة الأجهزة التي تصاب بعطل خلال سنة من تركيبها .

(٢) نسبة الأجهزة التي تصاب بعطل خلال السنة الثانية من تركيبها .

١٧ ـ فيما يلي التوزيع العمري النسبي لموظفي وزارة التربية الذين تبرعوا بدمائهم خلال الحملة الأخيرة لجمع الدم .

فئات العمر بالسنوات	١٥ ـ ١٩	٢٠ ـ ٢٤	٢٥ ـ ٢٩
النسبة المئوية	٪٧	٪١٨	٪٣٠

٣٠ ـ ٣٩	٤٠ ـ ٤٩	٥٠ ـ ٦٤	٦٥ ـ ٧٩	المجموع
٪٣٣	٪٧	٪٤	٪١	٪١٠٠

هل تتفق مع الرأي القائل بأن الموظفين في الفئة العمرية (٣٠ ـ ٣٩) هم أكثر الموظفين اقبالاً على التبرع بدمائهم ، يليهم الموظفون في الفئة العمرية (٢٥ ـ ٢٩) ؟ وضح سبب اجابتك .

١٨ ـ استخدمت الفئات : أقل من ١٠ ، ١٠ ـ ٢٩ ، ٣٠ ـ ٤٩ ، ٥٠ ـ ٩٩ ، ١٠٠ ـ ٢٩٩ ، ٣٠٠ أو أكثر ، لإنشاء التوزيع التكراري لعدد الجرائم التي وقعت في مختلف القرى والأحياء السكنية في الدولة خلال السنوات الخمس الماضية . لماذا نستخدم فئات غير متساوية في هذه الحالة ؟

١٩ ـ صنفت المهن إلى مهن مكتبية ومهن يدوية ومهن خدمات ورمز لهذه الأوجه بالرموز ١ ، ٢ ، ٣ على الترتيب ، كما صنف الأفراد إلى مدخنين (م) وغير مدخنين (غ) . فيما يلي بيانات عن المهنة والتدخين لمجموعة من ٤٩ شخصاً :

١، م	١، غ	٢، م	٢، غ	٣، غ	١، م	١، م	٢، غ
٢، غ	٣، م	٢، غ	١، م	٢، غ	٣، م	٢، غ	٣، غ
١، م	١، غ	١، غ	٢، م	١، م	٢، غ	١، م	١، م
٢، غ	٣، غ	٢، غ	٢، م	١، م	٢، م	٢، غ	٢، م
١، م	١، م	١، غ	٢، غ	١، غ	٣، م	٣، غ	١، م
٢، غ	١، م	١، م	٣، غ	٢، م	٢، غ	٣، م	

(أ) كون التوزيع التكراري المشترك للمهنة والتدخين في هذه البيانات .

(ب) هل هناك علاقة بين المهنة والتدخين ؟ اشرح سبب إجابتك .

٢٠ ـ يعطي الجدول التالي بيانات عن رأي مجموعة من الطلبة في الاقتراح الخاص بفتح المكتبة أيام الجمع والعطلات :

الفرقة الدراسية / رأي الطالب	الثانية	الثالثة	الرابعة	المجموع
موافق	٤٥	٣٠	٣٠	١٠٥
غير موافق	٣٠	١٥	٥٠	٩٥
المجموع	٧٥	٤٥	٨٠	٢٠٠

(أ) هل يختلف رأي الطلبة في الاقتراح باختلاف فرقهم الدراسية ؟ اشرح سبب إجابتك

(ب) كون التوزيع التكراري النسبي المشترك المناظر لهذا الجدول .

٢١ ـ يعطي الجدول الآتي التوزيع التكراري المشترك لعدد الأطفال في الأسرة وعدد المجلات الأسبوعية التي تشتريها الأسرة في مجموعة من ٦٠٠٠أسرة.

عدد المجلات / عدد الاطفال	صفر	١	٢	٣	٤	المجموع
صفر	٧٥٤	٨٦٨	٧٨١	١٦٢	٦٣	٢٦٢٨
١	٢٩٧	٣٦٣	٣٣٠	٨٧	٢٣	١١٠٠
٢	٢٥٦	٣٥٨	٣٠٧	٧١	٣٢	١٠٢٤
٣ أو أكثر	٣١٤	٣٩١	٣٨٢	٩٩	٦٢	١٢٤٨
المجموع	١٦٢١	١٩٨٠	١٨٠٠	٤١٩	١٨٠	٦٠٠٠

(أ) احسب التوزيع النسبي داخل كل عمود من أعمدة هذا الجدول.

(ب) هل تدل البيانات على وجود علاقة بين عدد أطفال الأسرة وعدد المجلات التي تشتريها الأسرة. وضح سبب الاجابة.

٢٢ ـ يعطي الجدول الآتي توزيع نفس الأسر في السؤال السابق حسب مستوى الدخل الشهري للأسرة وعدد المجلات الاسبوعية التي تشتريها الأسرة.

عدد المجلات / الدخل بالدراهم	صفر	١	٢	٣	٤	المجموع
أقل من ٤٠٠٠	١٠٠٥	٩٠٥	٤٥٢	١٢٦	٢٥	٢٥١٣
٤٠٠٠ ـ ٩٩٩٩	٥٣٦	٩٣٠	١١٣٩	٢٧٤	١٠٧	٢٩٨٦
١٠٠٠٠ أو أكثر	٨٠	١٤٥	٢٠٩	١٩	٤٨	٥٠١
المجموع	١٦٢١	١٩٨٠	١٨٠٠	٤١٩	١٨٠	٦٠٠٠

هل يدل هذا الجدول على وجود علاقة بين مستوى الدخل وعدد المجلات الأسبوعية التي تشتريها الأسرة؟ وضح سبب الإجابة.

٢٣ ـ فيما يلي درجات ٣٦ طالباً في اللغة العربية والاقتصاد :

اللغة العربية	الاقتصاد	اللغة العربية	الاقتصاد	اللغة العربية	الاقتصاد
٥٣	٧٠	٣١	٧٢	٣٥	٥٧
٢٤	٣٨	٨٠	٨٦	٥٦	٧٢
٦٢	٥٥	٤٥	٤٦	٦٥	٦٣
٩٠	٧٨	٧٨	٥٧	٧٨	٧٦
١٨	٣٥	٧١	٧١	٤٩	٥٣
٩٤	٩١	٨٤	٧٢	٨٢	١٠٠
٧٣	٦٩	٥٨	٥٩	٢٢	٣٨
٨٥	٨٣	٩	١٤	٩٠	٨٢
٢٥	٥١	١٦	٤٢	٧٧	٨٢
٧١	٥٣	٩٧	٩٣	٣٥	١٩
٨١	٦٠	٦٥	٦١	٥٢	٤٣
٥٢	٥٨	٤٢	٥٨	٩٣	٧٩

(أ) ضع هذه البيانات في توزيع تكراري مشترك باستخدام الفئات ١ ـ ٢٠، ٢١ ـ ٤٠ ، ٤١ ـ ٦٠، ٦١ ـ ٨٠ ، ٨١ ـ ١٠٠ لكل من المتغيرين .

(ب) هل هناك علاقة بين درجة الطالب في اللغة العربية ودرجته في الاقتصاد ؟ اشرح سبب إجابتك .

٢٤ ـ قامت إحدى الشركات الصناعية بإدخال نظام جديد للسلامة في مصانعها . فيما يلي بيانات عن عدد الحوادث خلال أسبوع قبل تركيب النظام وعدد الحوادث خلال أسبوع مناظر بعد تركيب النظام في ثلاثين مصنعاً من مصانع الشركة :

قبل	بعد	قبل	بعد	قبل	بعد
٣٧	٢٨	٧٣	٤٨	٢٧	٢٥
٧٢	٥٩	١١١	٨٣	٤٩	١٨
٢٦	٢٤	٤١	٣٨	٦٦	٦١

قبل	بعد	قبل	بعد	قبل	بعد
١٢٥	١٢٠	٧٧	٧٢	٨٣	٤٢
٤٥	٤٦	٧٣	٢٨	٦٣	٥٧
٥٤	٤٣	١٠٣	٦٩	٨٦	٥٠
١٣	١٥	٣٢	١٨	٤٥	٢٣
٧٩	٧٥	٦٨	٧٣	١٠٣	٦٧
٤٢	٢٣	٨٢	٦٥	٥٥	٥٧
٣٩	٣٥	٤٧	٤٢	٤١	٢٨

أ ـ ضع هذه البيانات في توزيع تكراري مشترك باستخدام الفئات ١٠ ـ ٣٩، ٤٠ ـ ٦٩، ٧٠ ـ ٩٩، ١٠٠ ـ ١٢٩ لكل من المتغيرين .

ب ـ احسب التوزيع التكراري النسبي المشترك .

٢٥ ـ يعطي الجدول التالي التوزيع التكراري المشترك لثلاث متغيرات هي الزمن (١٩٨٠ ـ ١٩٨٥) ، ونوع الجريمة (عنيفة ، غير عنيفة) والحكم في الجريمة (المتهم مذنب ، المتهم بريء) وذلك للجرائم المسجلة في دولة ما :

	١٩٨٠		١٩٨٥	
	جريمة عنيفة	جريمة غير عنيفة	جريمة عنيفة	جريمة غير عنيفة
المتهم مذنب	٢٥٠٠	١٠٠٠٠	٣٠٠٠	١٦٠٠٠
المتهم بريء	١٥٠٠	٤٠٠٠	٤٠٠٠	١١٠٠٠
المجموع	٤٠٠٠	١٤٠٠٠	٧٠٠٠	٢٧٠٠٠

أ ـ استكمل خلايا الجداول التالية :

(١)

السنوات / نوع الجريمة	١٩٨٠	١٩٨٥	المجموع
جريمة عنيفة			
جريمة غير عنيفة			
المجموع			

(٢)

نوع الحكم / السنوات	١٩٨٠ ١٩٨٥	المجموع
المتهم بذنب المتهم بريء		
المجموع		

(ب) هل اختلف نمط انواع الجرائم بين عامي ١٩٨٠ ، ١٩٨٥ ؟ اشرح سبب اجابتك .

(حـ) هل اختلف نمط أنواع الحكم في الجرائم بين عامي ١٩٨٠، ١٩٨٥ ؟ اشرح سبب إجابتك .

الباب الرابع

الرسوم البيانية

١ ـ مقدمة

تحدثنا في الباب السابق عن استخدام الجداول كأداة لعرض وتحليل البيانات الإحصائية ، وشرحنا كيف يمكن الاعتماد عليها في وصف نمط الاختلاف في متغير ما من ناحية ، وفي دراسة طبيعة العلاقات بين المتغيرات المختلفة من ناحية أخرى . ويمثل انشاء الجداول في معظم الحالات خطوة أولى على جانب كبير من الأهمية في التحليل الاحصائي للبيانات .

تستخدم الرسوم البيانية إلى جانب الجداول الاحصائية كأسلوب مكمل لها في شرح وتوضيح الحقائق الأساسية في البيانات . وتتميز الرسوم البيانية بأنها أشكال تعتمد على التأثير البصري لجذب اهتمام القارىء . هذا فضلاً عن عرض الاتجاهات الرئيسية في البيانات بشكل ميسر ، مع ما يترتب على ذلك من امكانيات استخدام هذه الرسوم عند عقد المقارنات بين مجموعات البيانات المختلفة .

وتؤدي الرسوم البيانية وظائف إحصائية متعددة أهمها ما يلي :

أ ـ عرض البيانات الاحصائية بأسلوب فعال : ذلك أن الشكل البياني الجيد يؤدي إلى إعطاء صورة صادقة وواضحة وجذابة للوضع المشاهد في البيانات . ويترتب على ذلك خلق انطباعات محددة لدى القارىء ، وهو أمر قد لا يكون من السهل تحقيقه بمجرد الاعتماد على الجداول بما فيها من أرقام صماء .

ب ـ الاستخدام كأداة من أدوات التحليل الاحصائي : يفيد الرسم البياني كأساس للتعرف على الشكل العام للنماذج والأساليب الاحصائية التي تصلح للاستخدام في تحليل البيانات ، ثم دراسة مدى كفاءة هذه النماذج في تمثيل البيانات فعلاً . وتستخدم الرسوم البيانية أيضاً في بعض الأحيان للتعبير عن المنحنيات المختلفة المشاهدة في البيانات وذلك عوضاً عن محاولة تمثيل هذه المنحنيات بصيغ رياضية معقدة . وتعتبر الرسوم البيانية الخاصة بالتوزيعات التكرارية وأشكال الانتشار للعلاقة بين متغيرين والرسوم البيانية للسلسلات الزمنية أمثلة على استخدامات الرسوم البيانية لأغراض التحليل الإحصائي .

ويذهب كثير من الكتاب إلى القول بضرورة أن تشمل عملية تحليل وتفسير أي مجموعة بيانات إحصائية على عرض بياني للمعلومات المتضمنة فيها . ويتطلب التخطيط السليم والتنفيذ الجيد للأشكال البيانية تحديد ومناقشة عدة أمور هي :

أ ـ الهدف من إنشاء الشكل البياني، ويشمل ذلك تحديد الحقائق التي يراد إبرازها أو وجهات النظر التي يراد تعضيدها . ويعتمد ذلك على دراسة البيانات المستخدمة بتأن وتحليلها بعناية .

ب ـ النوعية المتوقعة للقراء والمشاهدين للشكل البياني ، ويفيد ذلك في التعرف على الأسلوب الأمثل لمخاطبتهم .

حـ ـ الظروف التي يستخدم فيها الشكل البياني ، ويقصد بذلك ما إذا كان الشكل سوف يستخدم في تقرير مكتوب أو في محاضرة أو في برنامج تليفزيوني ، . . . الخ . هذا بالاضافة إلى التعرف على الأدوات الفنية المتاحة من أجهزة بصرية وشاشات يمكن أن تستخدم لعرض الشكل البياني . ويؤثر ذلك في تحديد الحجم المطلوب للشكل البياني والمواد المستخدمة في إعداده .

د ـ الأسلوب المستخدم لرسم البيانات ، وقد يكون ذلك خطاً أو منحنى بيانياً ، أو مجموعة من الأعمدة أو الدوائر ، . . . الخ . ويتم اختيار هذا الأسلوب في ضوء الاعتبارات التي تتحدد في (أ) ، (ب) ، (حـ) .

هـ ـ المواد الكتابية المستخدمة في الرسم ، ونمط الألوان التوضيحية وكتابة الخطوط ، وقد يستعان في هذا الصدد بآراء رسام أو خطاط محترف .

٢ ـ مكونات الشكل البياني

يتكون الشكل البياني من عدة مركبات هي :

أ ـ **المحورين الأفقي والرأسي** : ويسمى المحور الأفقي أحياناً محور س بينما يسمى المحور الرأسي محور ص . وينتج عن تقاطع هذين المحورين تقسيم ورقة الرسم البياني إلى أربعة أجزاء كما يبدو في شكل (١). تسمى نقطة تقاطع المحورين نقطة الأصل أو نقطة الصفر ، بحيث تمثل القيم الموجبة على المحور الأفقي إلى يمين هذه النقطة والقيم السالبة إلى يسارها . كذلك تقاس القيم الموجبة على المحور الرأسي أعلى هذه النقطة بينما تمثل القيم السالبة أسفلها .

ويعتبر المربع I في شكل (١) أكثر المربعات استخداماً في التطبيقات الإحصائية ، فمثلاً عند عرض بيانات عن سلسلة زمنية كما في شكل (٢)، يؤخذ الزمن على المحور الأفقي بينما تمثل قيمة الظاهرة على المحور الرأسي حيث يلاحظ أن القيم المستخدمة على كلا المحورين موجبة .

وتستخدم المربعات I ، IV في شكل (١) في الحالات التي تأخذ فيها الظاهرة الممثلة على المحور الرأسي قيماً موجبة وأخرى سالبة ، كما في شكل (٣) . وتستخدم المربعات I ، II في الحالات التي تكون فيها الظاهرة الممثلة على المحور الأفقي ذات قيم موجبة أو سالبة كما في شكل (٤) . وتجدر الإشارة إلى أنه نادراً ما يستخدم المربع III في التطبيقات الاحصائية لأن ذلك يتطلب أن قيم كل من الظاهرتين على المحور الأفقي والمحور الرأسي سالبة .

ص

+٥

+٤

المربع II　　+٣　　المربع I

+٢

+١

-٤　-٣　-٢　-١　　+١　+٢　+٣　+٤　س

-١

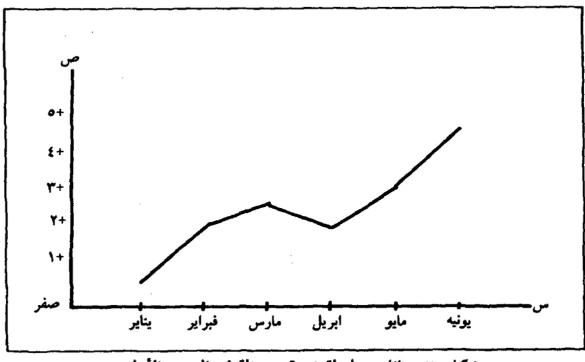
-٢

المربع III　　-٣　　المربع IV

-٤

-٥

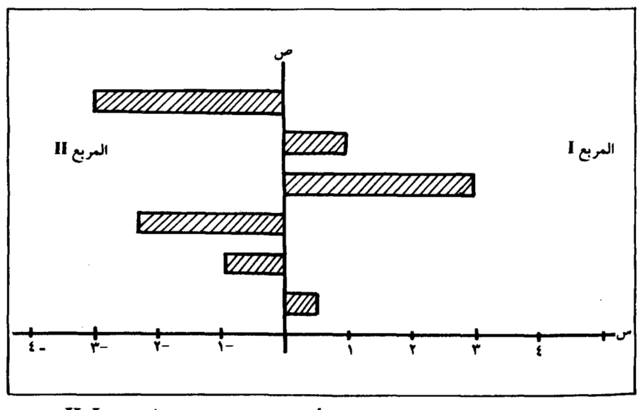
شكل (١) التقسيم بالمحاور



شكل (٢) بيانات سلسلة زمنية (ممثلة في المربع الأول)

ص
المربع I
+٣
+٢
+١
صفر
س
-١
-٢
المربع IV
-٣

شكل (٣): بيانات سلسلة زمنية تأخذ قيماً موجبة وأخرى سالبة (تمثل في المربعين I، IV)

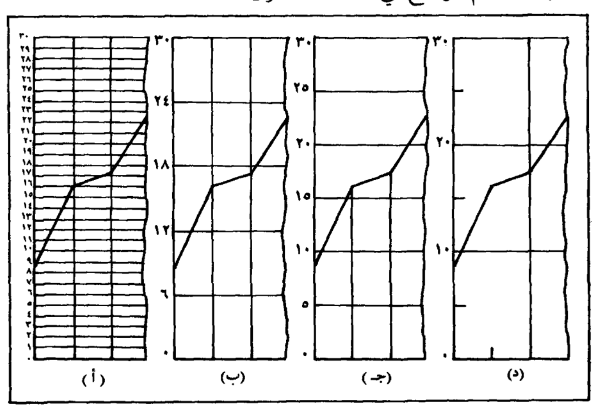


شكل (٤): أعمدة أفقية تمثل قيما سالبة وأخرى موجبة (استخدام المربعين I، II)

ب ـ عنوان الشكل البياني : يكون للشكل البياني عنواناً موجزاً وواضحاً يدل على محتواه . ويكتب هذا العنوان بخط بارز ويوضع عند مركز الشكل .

حـ ـ عناوين المحاور ووحدة القياس على كل محور : ويقصد بذلك كتابة إسم الظاهرة التي يمثلها كل محور سواء كان ذلك الزمن أو الدخل السنوي أو عدد السكان أو عدد الجرائم أو قيمة المبيعات السنوية . . الخ ، ثم كتابة وحدات قياس كل ظاهرة من هذه الظواهر فالزمن قد يكون بالسنوات والدخل السنوي بآلاف الدراهم وعدد السكان بالمليون نسمة ، . . . وهكذا . وينصح بكتابة هذه المعلومات أفقياً لتسهيل قراءتها .

ء ـ مقياس الرسم المستخدم على كل محور : يختار مقياس رسم على المحور الأفقي ومقياس رسم على المحور الرأسي بحيث تكون مساحة الشكل الناتج مناسبة لتحقيق الهدف من إنشائه . ويجب أن يقسم كل محور بشكل واضح تبعاً لمقياس الرسم المختار . ويوضح شكل (٥) طرقاً مختلفة لكتابة مقياس الرسم حيث يلاحظ أن أفضل هذه الطرق تظهر في الشكل (د) لأنها تتجنب الازدحام الواضح في الأشكال الأخرى .



شكل (٥) : طرق مختلفة لكتابة مقياس الرسم .

هـ ـ **جسم الشكل** : ويقصد بذلك المنحنيات أو الأعمدة المختلفة التي تظهر في الشكل البياني .

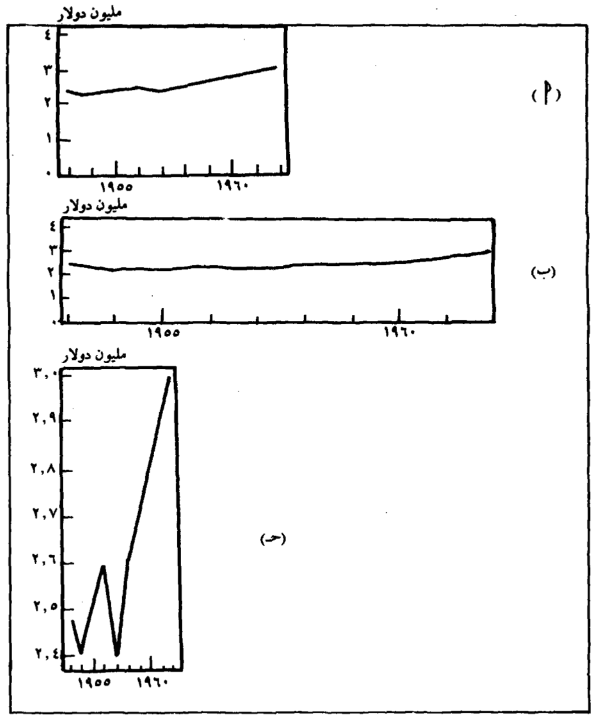
و ـ **مصدر البيانات** : يذكر مصدر البيانات التي تستخدم لانشاء الشكل البياني وذلك حتى يستطيع القارىء الرجوع إليه إذا دعت الحاجة . ويكتب هذا المصدر بخط صغير أسفل الشكل البياني .

ز ـ **ملاحظات توضيحية أخرى** : ويشمل ذلك كتابة عناوين واضحة لكل منحنى في الحالات التي يوجد فيها أكثر من منحنى في الشكل وشرح معنى أي رموز مستخدمة ووضع مفتاح مناسب للشكل إذا كان ذلك ضرورياً .

٣ ـ سوء استخدام الأشكال البيانية

قد يساء استخدام الأشكال البيانية عن عمد أو غير عمد . ويترتب على سوء الاستخدام إعطاء الانطباع الخاطىء عن الخصائص الأساسية للبيانات . وفيما يلي أهم العمليات التي قد تؤدي إلى أشكال بيانية مضللة .

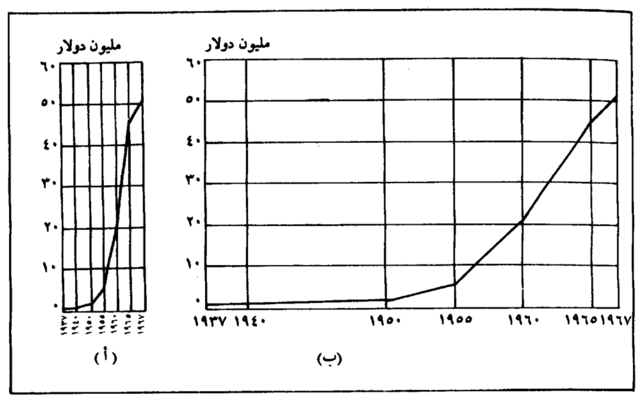
أ ـ **التلاعب في اختيار مقياس الرسم** : يمكن إظهار الاتجاه العام في البيانات على غير حقيقته بتوسيع أو تضييق مقياس الرسم على أي من المحورين . ويوضح شكل (٦) مثالاً لذلك حيث تعرض نفس مجموعة البيانات بثلاث طرق مختلفة . ويلاحظ وجود تناسب جيد بين مقياسي الرسم على المحورين في شكل ٦ (أ) ، ويخلص القارىء من دراسة هذا الشكل إلى وجود تغيرات طفيفة في الظاهرة التي تدرس . أما في شكل ٦ (ب) ، فقد انتفى التناسق بين مقياسي الرسم نتيجة تضييق المقياس على المحور الرأسي وتوسيعه على المحور الأفقي ، ويخرج القارىء نتيجة لذلك بانطباع أولي عن ثبات قيم الظاهرة التي تدرس . ويبدو تأثير تغيير مقياس الرسم بصورة اكثر حدة في شكل ٦ (جـ) حيث ترتب على تضييق مقياس الرسم على المحور الأفقي وتوسيعه على المحور الرأسي إلى إعطاء الانطباع الخاطىء بوجود تذبذبات واسعة في قيمة الظاهرة . ونخلص من ذلك إلى ضرورة دراسة



شكل (٦) : تأثير التلاعب في اختيار مقياس الرسم

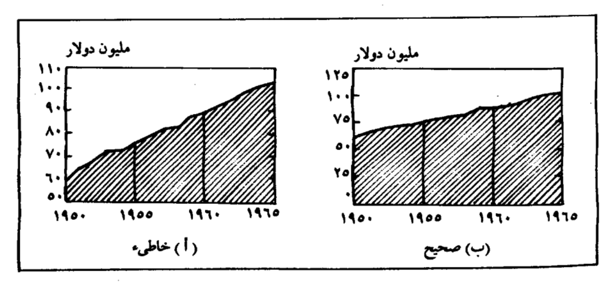
مقياس الرسم على كل من المحورين بعناية عند قراءة أو استخدام الأشكال البيانية .

ب ـ استخدام مقياس رسم خاطىء : يوضح شكل (٧) مثالاً لهذه الحالة ، حيث يلاحظ أن استخدام مقياس رسم خاطىء على المحور الأفقي في شكل ٧ (أ) يوحي بأن الارتفاع في الظاهرة التي تدرس كان ارتفاعاً حاداً وسريعاً ، على عكس حقيقة الوضع كما تظهر في شكل ٧ (ب) .

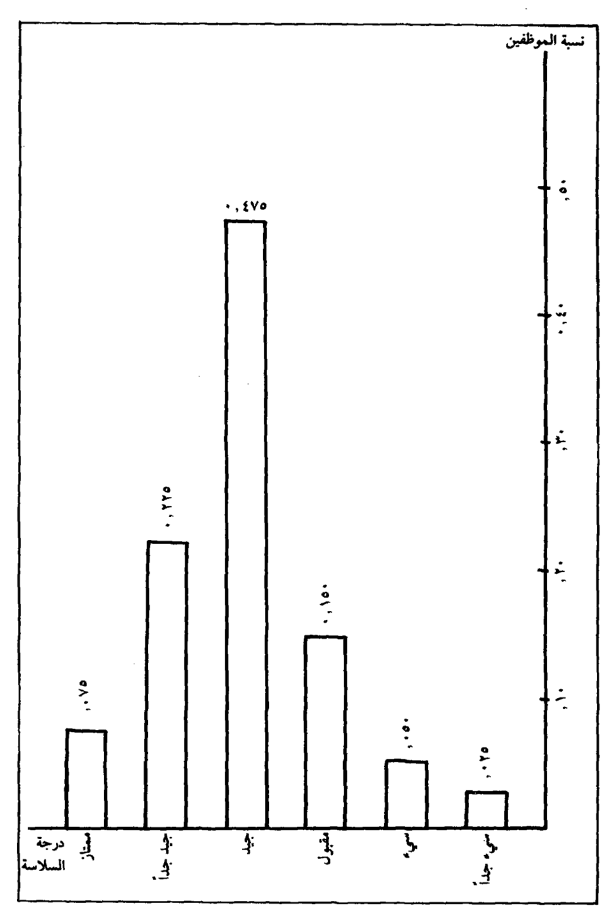


شكل (٧) : تأثير اختيار مقياس رسم خاطىء

جـ ـ عدم البدء عند الصفر على المحـور الرأسي : من الضـروري أن يبدأ المحور الـرأسي عند الصفـر حتى تتضح الصـورة الكاملة للتغيـرات في الظاهرة التي تدرس . ويجب على مستخدمي الرسوم البيانية التأكد من ذلـك عند شرح وتفسير هذه الرسوم . ويوضح شكل ٨ (أ) وشـكل ٨ (ب) التأثير البصري الخادع الذي ينتج من عدم البدء عند الصفر على المحور الرأسي .



شكل (٨) : تأثير عدم البدء عند الصفر على المحور الرأسي



شكل (٩): رأي السائقين حول درجة سلاسة قيادة سياراتهم في عينة من ٤٠ سائقاً.
(المصدر: جدول (٧) صفحة (٧١)).

٤ ـ العرض البياني للمتغيرات النوعية ـ طريقة الأعمدة

يعتبر أسلوب الأعمدة أكثر أساليب العرض ملاءمة لتمثيل البيانات النوعية . ويعتمد هذا الأسلوب على استخدام أعمدة رأسية أو أفقية لتمثيل التكرارات المناظرة للأوجه المختلفة للمتغير . فمثلاً يوضح شكل (٩) شكل الأعمدة المناظر للتوزيع التكراري النسبي لرأي السائقين في درجة سلاسة سياراتهم والذي يظهر في جدول (٧) صفحة (٧١) .

كمثال آخر ، يعطي جدول (١) التوزيع التكراري لتصاريح العمل الممنوحة في الدولة في عام ١٩٨٣ حسب الجنسية . ويظهر شكل الأعمدة المناظر في شكل (١٠)

جدول (١)

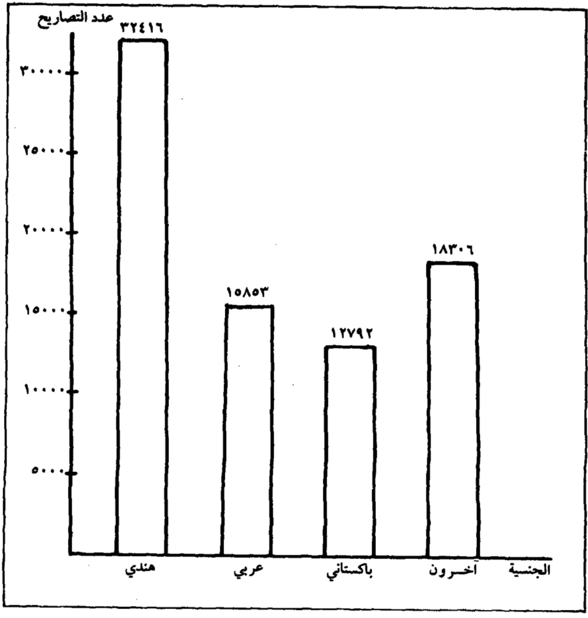
تصاريح العمل الممنوحة في الدولة عام ١٩٨٣ حسب الجنسية

الجنسية	عدد التصاريح
عربي	١٥٨٥٣
هندي	٣٢٤١٦
باكستاني	١٢٧٩٢
آخرون	١٨٣٠٦
المجموع	**٧٩٣٦٧**

(المصدر : جدول (٤٩) ـ المجموعة الاحصائية السنوية لدولة الامارات ـ وزارة التخطيط ، ١٩٨٤)

ويمكن إيجاز القواعد العامة لاستخدام أسلوب الأعمدة لعرض البيانات النوعية فيما يلي :

أ ـ يمثل كل تكرار في الجدول بمساحة العمود المناظر له . ويتحقق ذلك



شكل (١٠): تصاريح العمل الممنوحة في الدولة حسب الجنسية، ١٩٨٣
(المصدر : جدول (١)) .

بجعل قواعد الأعمدة متساوية في الطول ثم رسم كل عمود بحيث يكون ارتفاعه مناظراً للتكرار الذي يمثله .

ب ـ تترك مسافات منتظمة بين الأعمدة المختلفة لتسهيل التمييز بينها .

حـ ـ يجب أن تبدأ جميع الأعمدة عند خط واحد وأن يقاس طول كل منها من الصفر ، وذلك حتى يتضح النمط الكامل للاختلافات بينها .

ء ـ عند استخدام الأعمدة لعرض المتغيرات التصنيفية ، يفضل ترتيب هـذه

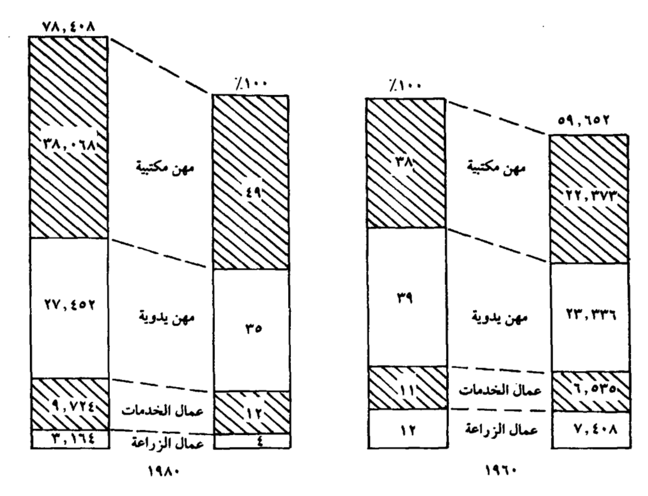
الأعمدة تصاعدياً أو تنازلياً لأن ذلك يساعد على توصيل المعلومات المتضمنة في الشكل بسرعة وسهولة .

هـ - يجب أن يحتوي الشكل البياني على مفتاح لشرح الرموز والألوان المستخدمة .

و - يكتب عنوان الشكل بوضوح ، ويذكر مصدر الحصول على البيانات المستخدمة في انشائه .

كمثال ثالث ، يوضح شكل (١١) استخدام أسلوب الأعمدة في أغراض المقارنة بين التوزيعات التكرارية النسبية . ويعتمد هذا الشكل على بيانات جدولي (٤)، (٥) صفحة (٦٩) عن التوزيع المهني للأشخاص في قوة العمل في بلد ما في عامي ١٩٦٠ ، ١٩٨٠ . رسم عمود يمثل العدد الكلي للأشخاص في قوة العمل في كل سنة من السنوات ، ثم جزىء هذا العمود حسب المهن المختلفة . ويظهر العمود الخاص بعام ١٩٨٠ في أقصى يسار الشكل بينما يظهر العمود الخاص بعام ١٩٦٠ في أقصى اليمين . رسم بعد ذلك عمودان يمثل كل منهما ١٠٠٪ من الأشخاص في قوة العمل في كل عام من العامين المذكورين ثم قسمت هذه الأعمدة حسب المهن المختلفة تبعاً للتكرار النسبي المئوي لكل مهنة . ويظهر هذان العمودان جنباً إلى جنب لتسهيل عملية المقارنة . فمثلاً يلاحظ أن نسبة العاملين في المهن اليدوية قد انخفضت من ٣٩٪ إلى ٣٥٪ وذلك على الرغم من أن العدد المطلق لهؤلاء العمال قد زاد من ٢٣٣٣٦ إلى ٢٧٤٥٢ .

وتفيد مثل هذه الأشكال في دراسة التغيرات التي تحدث في نصيب كل وجه من أوجه المتغير من عام إلى عام أو من مجتمع إلى آخر . مثال ذلك دراسة الاختلافات في نسبة العاملين في الدولة من جنسية معينة بين عام وآخر أو بين إمارة وأخرى ، كذلك دراسة التغيرات في نصيب نوع معين من السيارات في سوق السيارات بالدولة بين عام وآخر أو بين الطبقات الاجتماعية المختلفة .



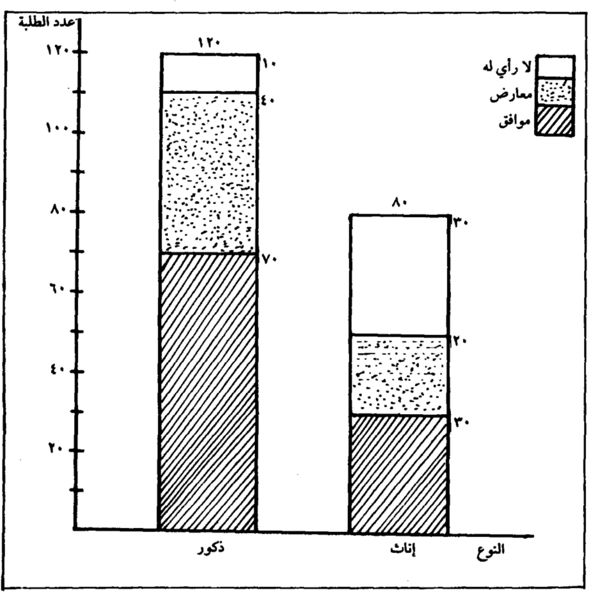
شكل (١١): التوزيع المهني المطلق والنسبي للأشخاص في قوة العمل في بلد ما في عامي ١٩٦٠ ، ١٩٨٠

جدول (٢)
التوزيع التكراري المشترك لنوع الطالب ولرأيه في اقتراح فتح المكتبة أيام الجمع لعينة من ٢٠٠ طالب

النوع / الرأي	ذكور	اناث	المجموع
موافق	٧٠	٣٠	١٠٠
معارض	٤٠	٢٠	٦٠
لا رأي له	١٠	٣٠	٤٠
المجموع	١٢٠	٨٠	٢٠٠

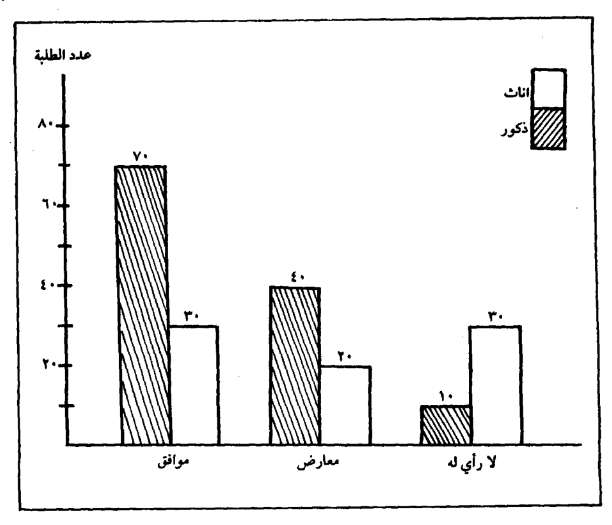
ويمكن أن يطور أسلوب الأعمدة ليستخدم في تمثيل التوزيع التكراري المشترك لمتغيرين نوعيين . فمثلاً ، يوضح جدول (٢) التوزيع التكراري المشترك لنوع الطالب ولرأيه في اقتراح فتح المكتبة أيام الجمع لعينة من ٢٠٠ طالب .

ويمكن عرض هذا الجدول بيانياً بأحد أسلوبين : أسلوب الأعمدة المجزأة ويظهر في شكل (١٢) وأسلوب الأعمدة المتلاصقة ويظهر في شكل (١٣) .



شكل (١٢): التوزيع التكراري المشترك لنوع الطالب ولرأيه في اقتراح فتح المكتبة أيام الجمع لعينة من ٢٠٠ طالب

(المصدر: جدول (٢)) .



شكل (١٣): التوزيع التكراري المشترك لنوع الطالب ولرأيه في اقتراح فتح المكتبة أيام الجمع لعينة من ٢٠٠ طالب
(المصدر : جدول (٢))

وترسم الأعمدة المجزأة بنفس الطريقة المعتادة ، حيث يوجد عمود مناظر لكل نوع (ذكور واناث) ثم يجزأ كل عمود الى ثلاثة أجزاء حسب الرأي . ويلاحظ أن هذا الأسلوب يسمح بإجراء المقارنة بين العدد الكلي في كل وجه من ناحية ثم دراسة التقسيم الداخلي لهذه الأعداد من ناحية أخرى . ويجب عند استخدام هذا الأسلوب أن يكون عدد اجزاء العمود صغيراً . وتجدر الملاحظة أن طريقة تجزئة الأعمدة تؤثر على وضوح عمليات المقارنة ، فمثلاً من السهل مقارنة بيانات الموافقين في شكل (١٢) لأن الأجزاء الخاصة بها تبدأ جميعاً على نفس الخط .

ويلاحظ عند استخدام أسلوب الأعمدة المتلاصقة ضرورة ترك مسافات منتظمة بين مجموعات الأعمدة ، كما يجب تمييز الأعمدة المختلفة بألوان ورموز واضحة . وتجدر الاشارة إلى أن هذا الاسلوب لا يسمح بالمقارنة بين الاعداد الكلية في الأوجه المختلفة بسهولة ووضوح .

٥ ـ العرض البياني للمتغيرات النوعية ـ طريقة الدائرة

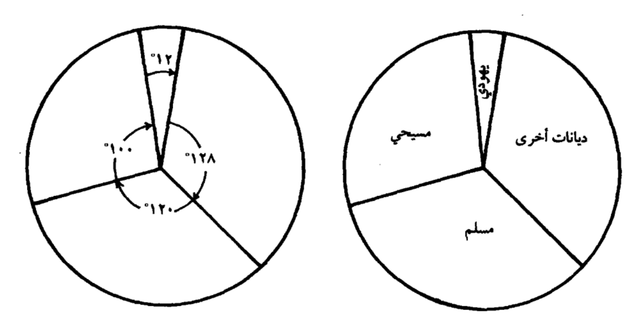
تفيد طريقة الدائرة في عرض البيانات التصنيفية ، حيث تقسم دائرة إلى قطاعات يمثل كل منها التكرار النسبي لأحد أوجه المتغير محل الدراسة . وترسم هذه القطاعات وفقاً للزوايا التي تتحدد درجاتها بضرب التكرارات النسبية في عدد درجات الزاوية الكلية بالدائرة وهو ٣٦٠° . ويوضح جدول (٣) كيفية حساب هذه الدرجات باستخدام بيانات عن توزيع الديانة في عينة عشوائية حجمها ١٨٠ شخصاً من سكان إحدى المدن (بيانات افتراضية) .

جدول (٣)

التوزيع التكراري للديانة

الديانة	عدد الأشخاص	التكرار النسبي	درجات زاوية القطاع
مسلم	٦٠	$\frac{٦٠}{١٨٠} = \frac{١}{٣}$	$\frac{١}{٣} \times ٣٦٠ = ١٢٠°$
مسيحي	٥٠	$\frac{٥٠}{١٨٠} = \frac{٥}{١٨}$	$\frac{٥}{١٨} \times ٣٦٠ = ١٠٠°$
يهودي	٦	$\frac{٦}{١٨٠} = \frac{١}{٣٠}$	$\frac{١}{٣٠} \times ٣٦٠ = ١٢°$
ديانات أخرى	٦٤	$\frac{٦٤}{١٨٠} = \frac{١٥}{٤٥}$	$\frac{١٥}{٤٥} \times ٣٦٠ = ١٢٨°$
المجموع	١٨٠	١	٣٦٠°

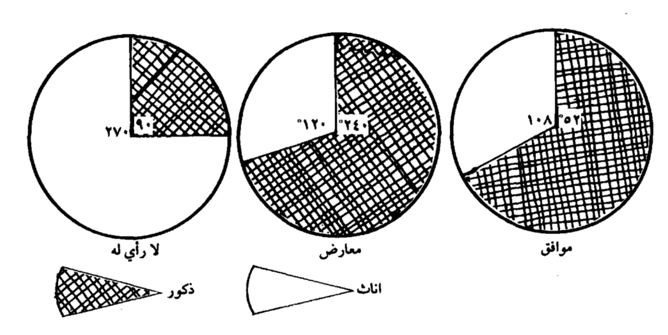
وترسم هذه الزوايا في الدائرة بالبدء من نصف القطر الرأسي (المناظر للساعة ١٢,٠٠) والاستمرار في اتجاه عقارب الساعة . ويحسن ترتيب القطاعات تنازلياً عند الرسم بدءاً بأكبر قطاع . ويظهر شكل الدائرة في شكل (١٤) .



شكل (١٤) : توزيع الديانة في عينة عشوائية حجمها ١٨٠ شخصاً من سكان إحدى المدن .

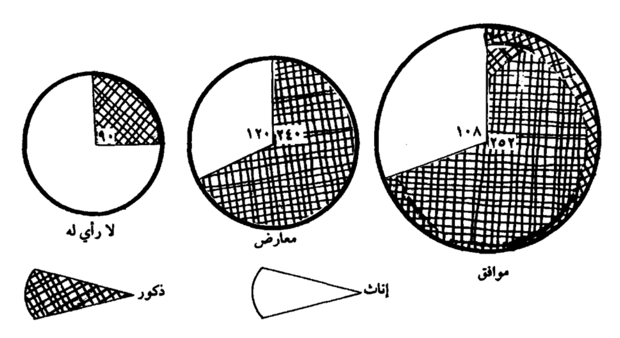
يتضح مما سبق أن شكل الدائرة يستخدم لعرض التركيبة النسبية لأوجه متغير ما . ويجب أن يكون عدد القطاعات في الدائرة قليلاً بقدر الامكان ، للمحافظة على وضوح الشكل . ويلاحظ كقاعدة عامة أن شكل الدائرة يكون أقل وضوحاً من شكل الأعمدة وذلك لصعوبة اجراء المقارنات بين مساحات قطاعات الدائرة بمجرد النظر .

ويمكن أن تستخدم طريقة الدائرة لعرض الجداول التكرارية المزدوجة . يعطي شكل (١٥) شكل الدائرة المناظر لبيانات جدول (٢) التي تمثل التوزيع المشترك للرأي والنوع في عينة من ٢٠٠ طالب . وقد رسمت الدوائر الثلاث بالطريقة المعتادة حيث حسبت التكرارات النسبية للذكور والاناث داخل كل وجه من أوجه الرأي ثم ضربت هذه النسب في ٣٦٠° . ويلاحظ أن الدوائر المرسومة في شكل (١٥) متساوية المساحة لأن كل منها يمثل مجموعاً كلياً قدره ١٠٠٪ هو مجموع التكرارات النسبية المئوية في كل حالة .

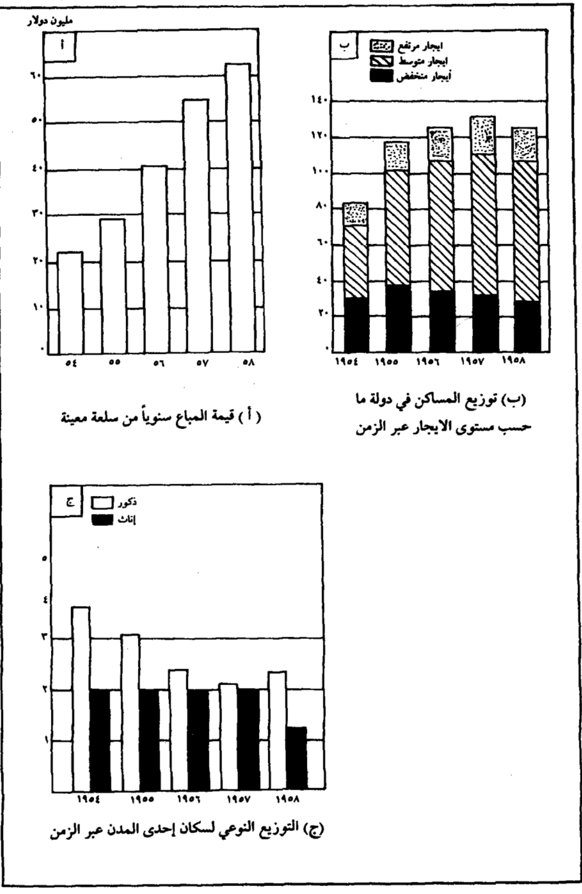


شكل (١٥): التوزيع النسبي لنوع الطالب لكل رأي من الآراء المتعلقة بفتح المكتبة أيام الجمع
(المصدر : جدول (٢))

ويمكن إعادة رسم شكل (١٥) لعرض الفروق بين التكرارات المطلقة . وفي هذه الحالة يتعين رسم دوائر مختلفة المساحة تمثل كل منها العدد المطلق الكلي المناظر لها . ويعطي شكل (١٦) الشكل الناتج في هذه الحالة .



شكل (١٦): التوزيع المشترك لنوع الطالب ولرأيه في اقتراح فتح المكتبة أيام الجمع لعينة من ٢٠٠ طالب



(أ) قيمة المباع سنوياً من سلعة معينة

(ب) توزيع المساكن في دولة ما حسب مستوى الايجار عبر الزمن

(ج) التوزيع النوعي لسكان إحدى المدن عبر الزمن

شكل (١٧): أمثلة لعرض السلسلات الزمنية بأسلوب الأعمدة .

وينصح كقاعدة عامة بعدم استخدام أسلوب الدوائر لعرض الاختلافات بين التكرارات المطلقة وذلك لصعوبة قياس وتفسير الفروق بين مساحات الدوائر أو بين مساحات القطاعات المختلفة في هذه الدوائر بمجرد النظر . ويجب تبعاً لذلك عدم التوسع في استخدام أسلوب الدائرة والاعتماد عليه فقط عند عرض الاختلافات النسبية في الظواهر .

٦ ـ العرض البياني للسلسلات الزمنية ـ الخط البياني

يمكن عرض السلسلات الزمنية بيانياً باستخدام أسلوب الأعمدة . ويوضح شكل (١٧) بعض أمثلة هذه الاستخدامات .

يلاحظ أن قمم الأعمدة تعكس التحركات التي تحدث في السلسلة . ويمكن تبعاً لذلك توصيل هذه القمم بخطوط كأسلوب بديل لعرض السلسلات الزمنية بيانياً . ويسمى الشكل الناتج في هذه الحالة الخط البياني . ويفيد استخدام الخط البياني كبديل للأعمدة في عدة مواقف أهمها :

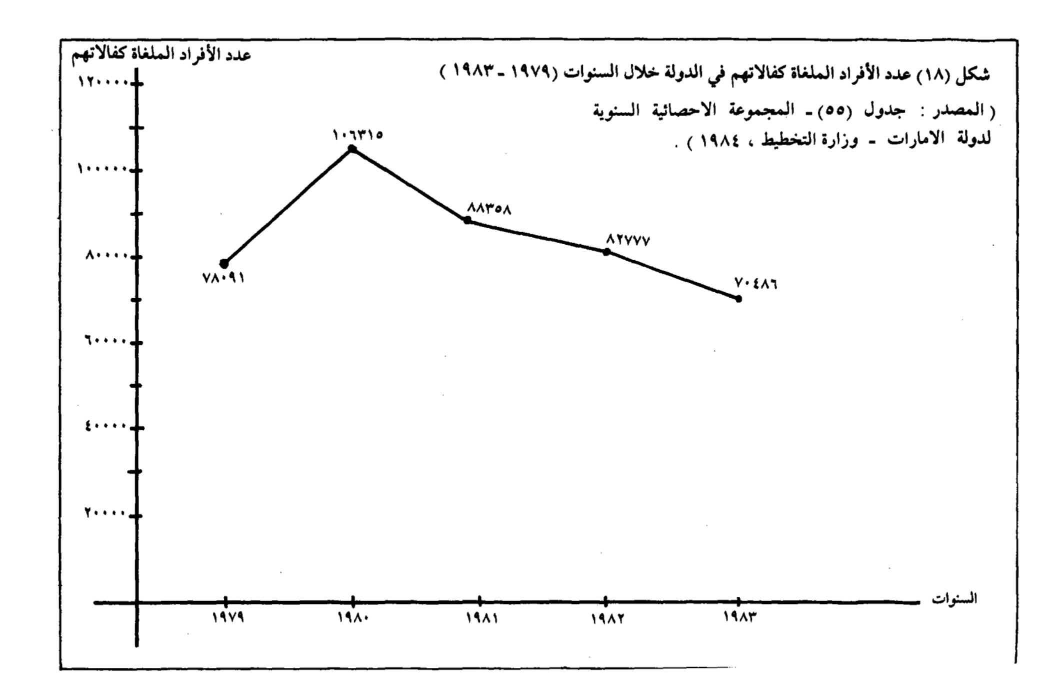
أ ـ عندما تكون السلسلة معطاة لفترة زمنية طويلة ، إذ أن استخدام الأعمدة في هذه الحالة يؤدي الى ازدحام الشكل وعدم وضوحه .

ب ـ عندما تكون السلسلة معطاة على فترات غير منتظمة ، أو تغطي فئات زمنية غير متساوية ، إذ أن استخدام الأعمدة في هذه الحالة قد يؤدي الى إعطاء صورة مضللة عن الاتجاه العام في البيانات .

حـ ـ عندما يكون الهدف هو اظهار الاتجاه العام في البيانات مع الزمن ، وبيان نمط التذبذبات المشاهد في البيانات .

ء ـ عندما يراد استخدام الشكل البياني كأداة للتقدير أو التنبؤ بقيمة الظاهرة في المستقبل ، وذلك بناء على الاتجاه العام المشاهد في البيانات .

هـ ـ عندما يكون من الضروري عرض أكثر من سلسلة زمنية في الشكل البياني الواحد ، إذ أن استخدام الخطوط البيانية في هذه الحالة يؤدي الى شكل بياني أقل ازدحاماً واكثر وضوحاً .



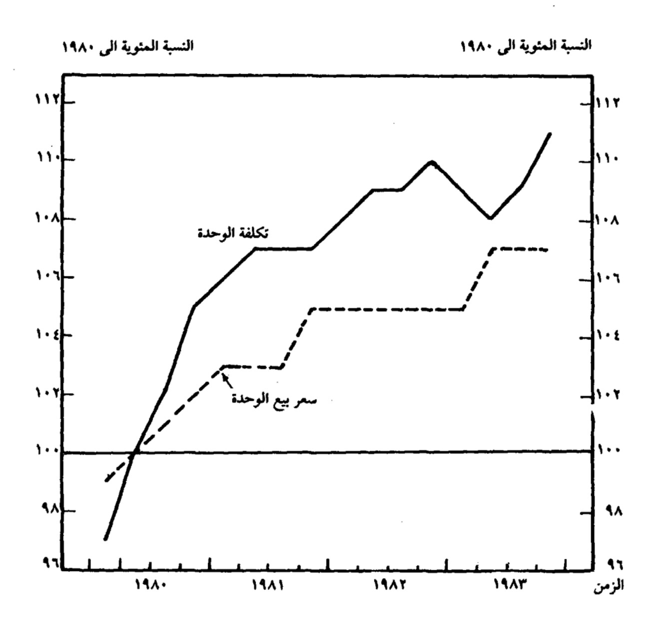
شكل (١٨) عدد الأفراد الملغاة كفالاتهم في الدولة خلال السنوات (١٩٧٩ ـ ١٩٨٣)

(المصدر : جدول (٥٥) ـ المجموعة الاحصائية السنوية لدولة الامارات ـ وزارة التخطيط ، ١٩٨٤) .

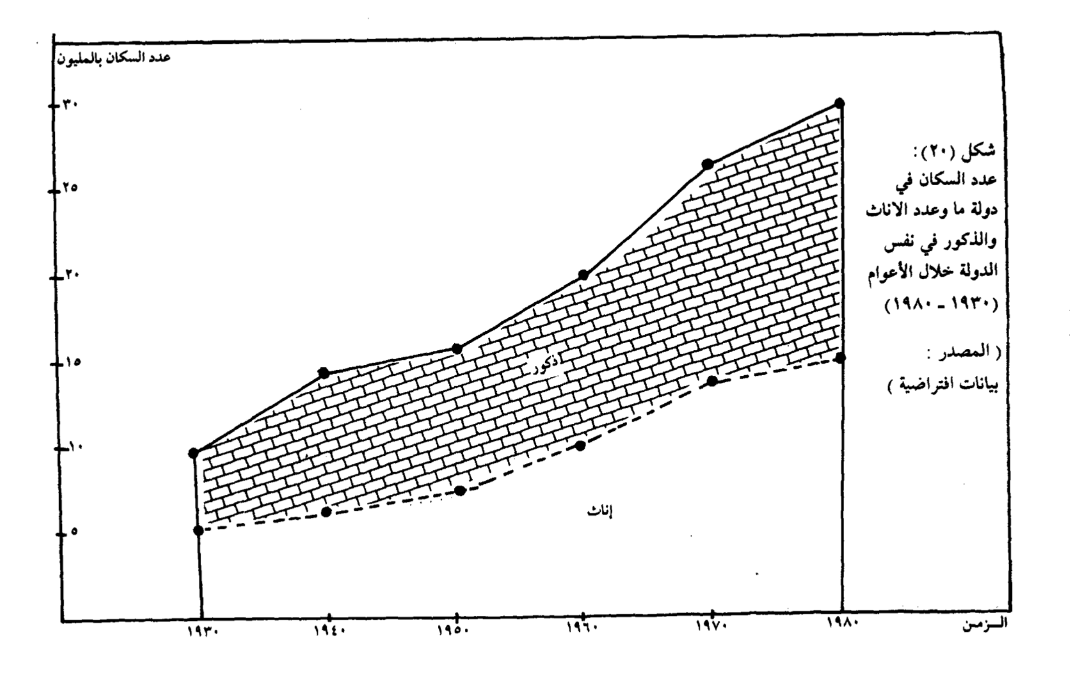
يعطي شكل (١٨) مثالاً لسلسلة زمنية عرضت بخط بياني. وتمثل هذه السلسلة عدد الأفراد الملغاة كفالاتهم سنوياً في الدولة خلال السنوات (١٩٧٩ ـ ١٩٨٣) . ويلاحظ أن المحور الأفقي يمثل الزمن بالسنوات بينما تمثل الظاهرة (عدد الأفراد الملغاة كفالاتهم) على المحور الرأسي .ويرسم الخط البياني بوضع نقطة عند كل سنة تناظر عدد الكفالات الملغاة في تلك السنة فمثلاً عند سنة ١٩٨٠ توضح نقطة على بعد عمودي قدره ١٠٦٣١٥ . توصل هذه النقط بخطوط مستقيمة للحصول على الخط البياني للسلسلة الزمنية .

وتوضح أشكال (١٩) ، (٢٠) أمثلة اخرى لخطوط بيانية . ويلاحظ من هذه الأشكال امكانية تطوير أسلوب الرسم للوفاء باحتياجات مختلفة .

يوضح شكل (١٩) امكانية استخدام الخط البياني لدراسة التغيرات التي تحدث في سلسلة زمنية من فترة لأخرى . يعطي الشكل تكلفة الوحدة المنتجة من سلعة ما وسعر بيع الوحدة من هذه السلعة مقاسة كنسبة مئوية من قيمها في عام ١٩٨٠ (أي أن عام ١٩٨٠ = ١٠٠) . ويلاحظ وجود خط أفقي عند ١٠٠ يمثل أساس المقارنة . ويتضح من الشكل أن التغير النسبي في التكلفة خلال (١٩٨٠ ـ ١٩٨٣) كان اكبر من التغير النسبي في سعر البيع خلال نفس الفترة . ويوضح شكل (٢٠) كيفية استخدام أسلوب الخط البياني في المواقف التي تستخدم فيها الأعمدة المجزأة . يعطي الشكل العدد الكلي للسكان في دولة ما عبر الزمن مع تقسيم هذا العدد حسب النوع . ويمكن الاعتماد على مثل هذا الشكل في دراسة التغيرات في العدد الكلي لظاهرة ما بالاضافة الى دراسة التغيرات في مكونات هذا العدد .

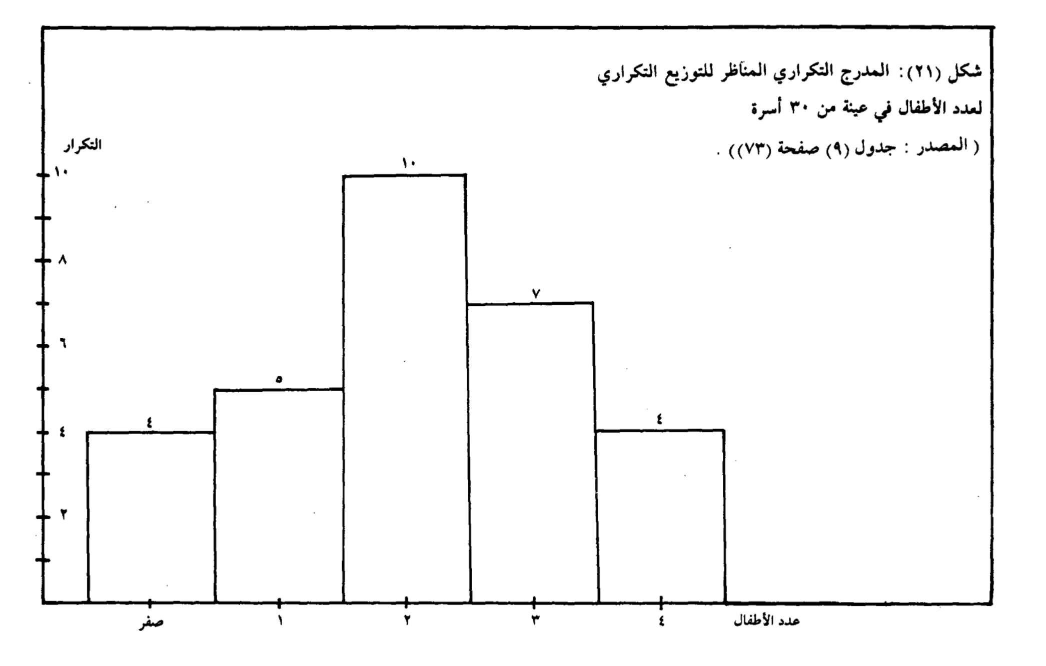


شكل (١٩): تكلفة الوحدة المنتجة من سلعة ما وسعر بيع الوحدة من هذه السلعة مقاساً كنسبة مئوية من قيمها في عام ١٩٨٠
(المصدر : بيانات افتراضية)



شكل (٢٠): عدد السكان في دولة ما وعدد الاناث والذكور في نفس الدولة خلال الأعوام (١٩٣٠ - ١٩٨٠)

(المصدر : بيانات افتراضية)



شكل (٢١): المدرج التكراري المناظر للتوزيع التكراري
لعدد الأطفال في عينة من ٣٠ أسرة
(المصدر : جدول (٩) صفحة (٧٣)) .

٧ ـ العرض البياني للمتغيرات الكمية ـ المدرج التكراري

ينشأ المدرج التكراري باستخدام طريقة الأعمدة لعرض التوزيع التكراري لمتغير كمي . فمثلاً ، يوضح شكل (٢١) المدرج التكراري المناظر للتوزيع التكراري لعدد الأطفال في الأسرة المعطى في جدول (٩) صفحة (٧٣). وقد تم رسم هذا المدرج بتمثيل القيم المختلفة للمتغير (وهو عدد الأطفال) على المحور الأفقي ، ثم رسم أعمدة تتمركز عند هذه القيم وتمثل التكرارات المناظرة . وتكون الأعمدة الناتجة متلاصقة .

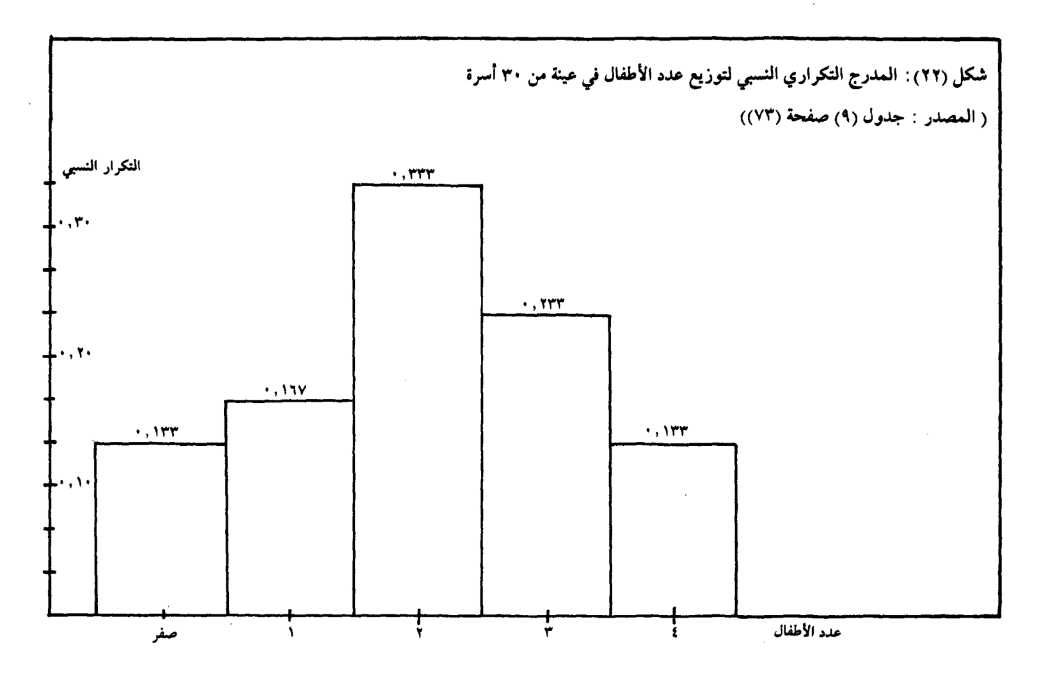
ويجب ان تكون مساحة كل عمود ممثلة للتكرار المناظر . ويتحقق ذلك برسم أعمدة تتناسب ارتفاعاتها مع التكرارات المناظرة طالما ان قواعد هذه الأعمدة تكون متساوية الطول . وتمثل المساحة الكلية للمدرج التكراري العدد الكلي للتكرارات .

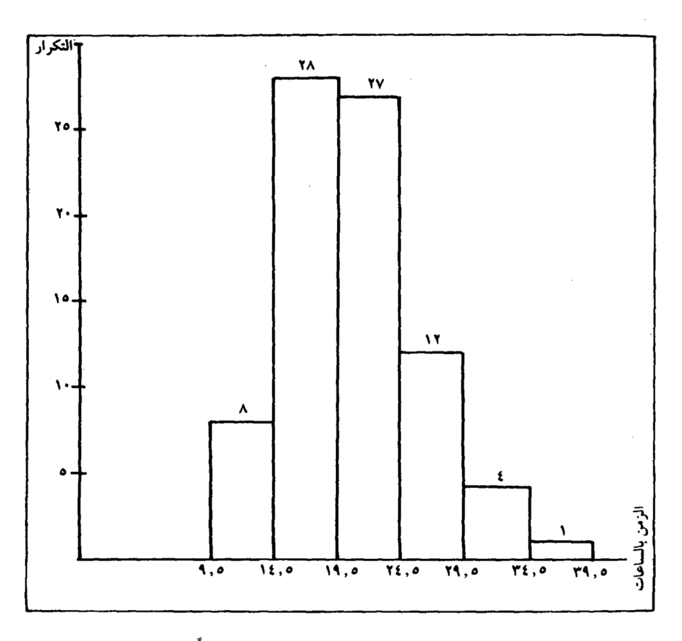
ويمكن أن يستخدم المدرج التكراري لعرض التكرارات النسبية . ويعطي شكل (٢٢) مدرج التكرارات النسبية لنفس بيانات عدد الأطفال في الأسرة . ويلاحظ ان المساحة الكلية تحت هذا المدرج تمثل الواحد ، وهو مجموع التكرارات النسبية . ويلاحظ كذلك ان نمط التوزيع النسبي مطابق تماماً لنمط التوزيع المطلق في شكل (٢١) ، وينحصر الاختلاف بين الشكلين في مقياس الرسم المستخدم على المحور الرأسي فقط .

يعطي شكل (٢٣) المدرج التكراري المناظر لبيانات عدد الساعات التي يقضيها الطلبة في ممارسة هواياتهم والمعطاة في جدول (١٦) صفحة (٨٦)، وذلك كمثال على مدرج تكراري لمتغير متصل. ويلاحظ ان عدد الساعات يظهر على المحور الأفقي . وتمثل كل فئة ببعد متساو على هذا المحور ، حيث ان الفئات متساوية الطول . وينشأ المدرج التكراري برسم عمود على كل فئة يتناسب ارتفاعه مع التكرار المناظر لهذه الفئة . ويترتب على ذلك استيفاء الخاصية الأساسية للمدرج التكراري والتي تتطلب ان تكون

شكل (٢٢) : المدرج التكراري النسبي لتوزيع عدد الأطفال في عينة من ٣٠ أسرة

(المصدر : جدول (٩) صفحة (٧٣))

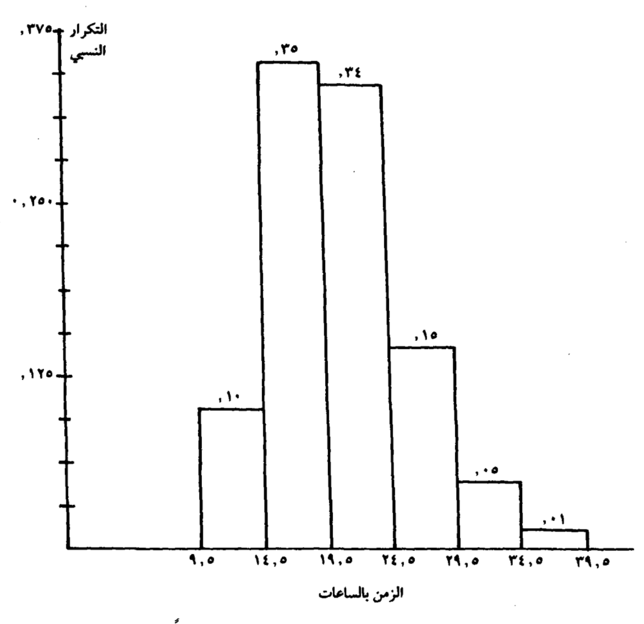




شكل (٢٣): المدرج التكراري لتوزيع مفردات عينة من ٨٠ طالباً حسب عدد الساعات التي يقضيها كل منهم في ممارسة هواياته اسبوعياً
(المصدر : جدول (١٦) صفحة (٨٦)) .

مساحة كل عمود ممثلة للتكرار المناظر ، وان تكون المساحة الكلية للمدرج ممثلة للمجموع الكلي للتكراري .

ويمكن عرض بيانات نفس التوزيع باستخدام التكرارات النسبية كما يظهر في شكل (٢٤) ، حيث يلاحظ ان المساحة الكلية تحت المدرج التكراري تساوي الواحد .



شكل (٢٤): التوزيع التكراري النسبي لمفردات عينة من ٨٠ طالباً حسب عدد الساعات التي يقضيها كل منهم في ممارسة هواياته اسبوعياً
(المصدر : جدول (١٦) صفحة (٨٦))

قد تكون فئات التوزيع التكراري غير متساوية الطول ، ويعني ذلك ان قواعد الأعمدة التي تناظر هذه الفئات تكون ايضاً غير متساوية . ويجب ان يراعى ذلك عند رسم الأعمدة بحيث تكون مساحة كل عمود (وليس ارتفاعه) ممثلة للتكرار المناظر . ولما كانت مساحة العمود = القاعدة × الارتفاع ، فإنه ينبغي للحصول على المساحات الصحيحة أن ترسم الأعمدة بحيث تحقق ارتفاعاتها العلاقة :

$$\text{ارتفاع العمود} = \frac{\text{التكرار المناظر}}{\text{طول الفئة المناظرة}}$$

وتسمى ارتفاعات الأعمدة في هذه الحالة أحياناً بالتكرارات المعدلة . ويتم حساب التكرار المعدل لكل فئة من فئات الجدول قبل القيام برسم المدرج التكراري . ويتضح ذلك من المثال التالي . يعطي جدول (٤) التوزيع العمري للمرضى الذين عولجوا في إحدى المستشفيات خلال أسبوع معين . ولما كانت فئات التوزيع غير متساوية الطول فإنه من الضروري تعديل التكرارات . ويوضح جدول (٥) خطوات حساب التكرارات المعدلة في هذه الحالة .

جدول (٤)

التوزيع العمري للمرضى الذين عولجوا في احدى المستشفيات خلال أسبوع معين

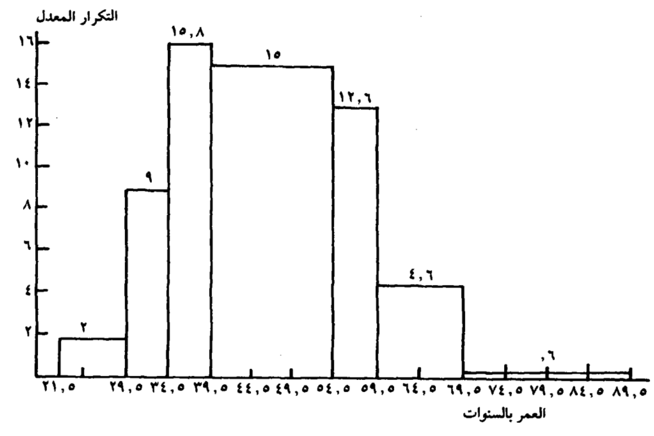
العمر لأقرب سنة	٢٢ ـ ٢٩	٣٠ ـ ٣٤	٣٥ ـ ٣٩	٤٠ ـ ٥٤	٥٥ ـ ٥٩	٦٠ ـ ٦٩	٧٠ ـ ٨٩	المجموع
عدد المرضى	١٦	٤٥	٧٩	٢٢٥	٦٣	٤٦	١٢	٤٨٦

جدول (٥)

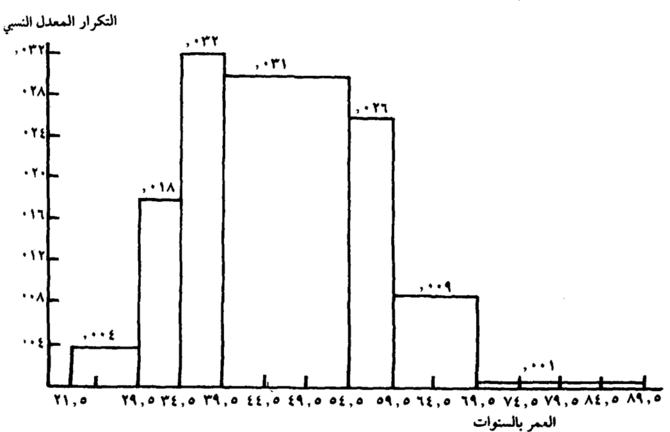
حساب ارتفاعات الأعمدة (التكرارات المعدلة) اللازمة لرسم المدرج التكراري للتوزيع العمري للمرضى

فئة العمر	عدد المرضى	طول الفئة	التكرار المعدل (ارتفاع العمود)
٢٢ ـ ٢٩	١٦	٨	١٦ ÷ ٨ = ٢
٣٠ ـ ٣٤	٤٥	٥	٤٥ ÷ ٥ = ٩
٣٥ ـ ٣٩	٧٩	٥	٧٩ ÷ ٥ = ١٥,٨
٤٠ ـ ٥٤	٢٢٥	١٥	٢٢٥ ÷ ١٥ = ١٥
٥٥ ـ ٥٩	٦٣	٥	٦٣ ÷ ٥ = ١٢,٦
٦٠ ـ ٦٩	٤٦	١٠	٤٦ ÷ ١٠ = ٤,٦
٧٠ ـ ٨٩	١٢	٢٠	١٢ ÷ ٢٠ = ,٦

(المصدر : جدول (٤))



شكل (٢٥): التوزيع العمري للمرضى الذين عولجوا في احدى المستشفيات خلال أسبوع معين



شكل (٢٦) : التوزيع العمري النسبي للمرضى الذين عولجوا
في أحد المستشفيات خلال أسبوع معين

ويظهر المدرج التكراري المناظر في شكل (٢٥) ، حيث رسمت الأعمدة بحيث تكون ارتفاعاتها ممثلة للتكرارات المعدلة المناظرة . ويمكن للقارىء ان يتأكد من ان مساحات هذه الأعمدة تمثل التكرارات الفعلية المناظرة كما هو مطلوب في المدرج التكراري . ويجب التأكيد على أن رسم المدرجات التكرارية ذات الفئات غير متساوية الطول دون تعديل التكرارات هو أمر خاطىء ويؤدي الى اعطاء نمط مضلل للتوزيع .

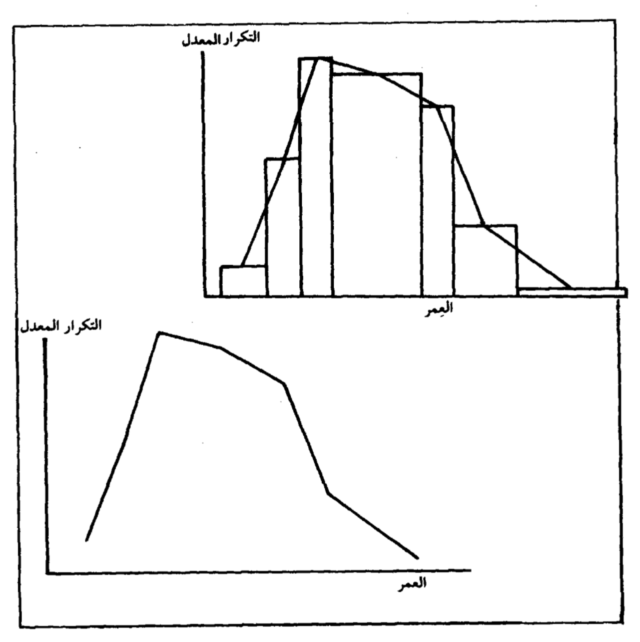
ويمكن الاستعاضة عن شكل (٢٥) برسم مدرج للتكرارات النسبية ، ويظهر هذا المدرج في شكل (٢٦) . ويمكن للقارىء ان يتأكد من ان المساحة الكلية لهذا المدرج تساوي الواحد .

٨ ـ العرض البياني للمتغيرات الكمية ـ المضلع التكراري

المضلع التكراري هو خط بياني ينشأ بالتخلص من الأعمدة في المدرج التكراري بعد توصيل مراكز قمم هذه الأعمدة بخطوط مستقيمة . ويوضح شكل (٢٧) المضلع التكراري المناظر للتوزيع العمري للمرضى المعطى في جدول (٤) .

ويمكن رسم المضلع التكراري مباشرة بتمثيل التكرارات في الفئات المختلفة بنقط توضع عند مراكز هذه الفئات ، ثم التوصيل بين هذه النقط بخطوط مستقيمة . ويلاحظ ان العلاقة بين المدرج التكراري والمضلع التكراري تعني انه من الضروري تعديل التكرارات عند رسم المضلع التكراري لتوزيع ذو فئات غير متساوية الطول وان المساحة الكلية تحت المضلع التكراري تساوي الواحد تقريباً .

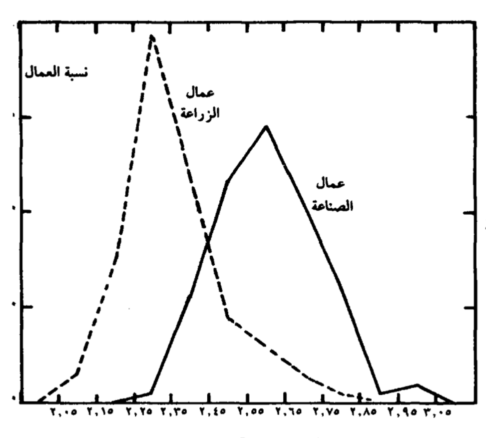
ويفيد كل من المدرج التكراري والمضلع التكراري في اعطاء وصف سريع وبسيط لنمط الاختلاف في البيانات . اذ يمكن بالنظر الى الشكل تحديد المدى الذي ينتشر فوقه التوزيع ، والأهمية النسبية لقيم وفئات المتغير بالاضافة الى المركز التقريبي للتوزيع ، ونمط الانتشار حول هذا المركز .



شكل (٢٧): المضلع التكراري للتوزيع العمري للمرضى الذين عولجوا في أحد المستشفيات خلال أسبوع معين

ويفضل استخدام المضلع التكراري عندما يكون عدد فئات التوزيع كبيراً او عندما يراد اجراء مقارنة بين توزيعات تكرارية مختلفة .

يوضح شكل (٢٨) مثالاً لاستخدام المضلع التكراري لأغراض المقارنة ، حيث يظهر المضلع التكراري لأجور مجموعة من عمال الصناعة والمضلع التكراري لأجور مجموعة من عمال الزراعة . وتبين المقارنة ان عمال الصناعة يحصلون بشكل عام على أجور أعلى من تلك التي يحصل عليها عمال الزراعة ، كذلك يلاحظ ان أجور عمال الزراعة اكثر تركزاً حول مركز التوزيع .



شكل (٢٨): المضلع التكراري النسبي لأجور مجموعة من عمال الزراعة والمضلع التكراري النسبي لأجور مجموعة من عمال الصناعة - (بيانات افتراضية)

التكرار التجميعي الصاعد
٨٠ ٧٠ ٦٠ ٥٠ ٤٠ ٣٠ ٢٠ ١٠
٨ ٣٦ ٦٣ ٧٥ ٧٩ ٨٠
١٠ ١٥ ٢٠ ٢٥ ٣٠ ٣٥ ٤٠
عدد الساعات

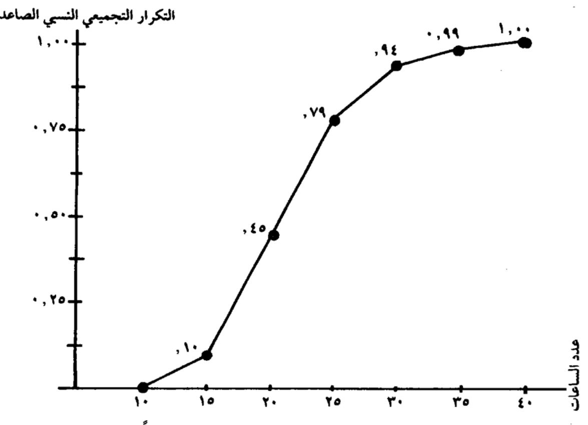
شكل (٢٩): المضلع التجميعي الصاعد لعينة من ٨٠ طالبا حسب عدد الساعات التي يقضونها في ممارسة هواياتهم .

٩ ـ العرض البياني للمتغيرات الكمية ـ المضلعات التجميعية

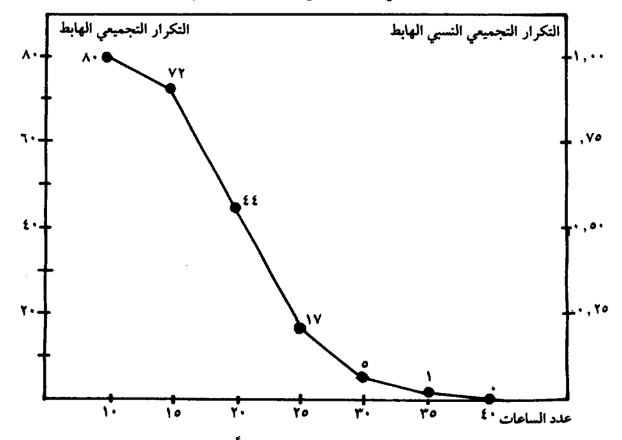
يفيد المضلع التجميعي في كثير من الاستخدامات الاحصائية ، وينشأ برسم خط بياني لتمثيل التوزيع التكراري التجميعي . يعطي شكل (٢٩) المضلع التجميعي الصاعد المناظر لجدول (١٨) صفحة (٨٩). وقد رسم هذا المضلع بأخذ قيم المتغير (وهو عدد الساعات التي يقضيها الطالب في ممارسة هواياته) على المحور الأفقي وتمثيل التكرارات التجميعية الصاعدة المناظرة على المحور الرأسي . وتوضع نقطة عند كل قيمة من قيم المتغير تمثل التكرار التجميعي المناظر . ويلاحظ ان وضع النقط عند القيم تماماً أمر منطقي نظراً لأن التكرارات يتم تجميعها حتى هذه القيم . ويتم التوصيل بين جميع هذه النقط بخطوط مستقيمة للحصول على المضلع التجميعي الصاعد .

يستخدم المضلع التجميعي الصاعد في شكل (٢٩) لوصف نمط الاختلاف في البيانات . فمثلاً يلاحظ ان ٤٠ طالباً يقضون أقل من ٢١ ساعة أسبوعياً في ممارسة هواياتهم وان ٢٠ طالباً يقضون أقل من ١٨ ساعة وهكذا . وقد يتضح نمط الاختلاف بصورة أفضل اذا رسمت التكرارات التجميعية النسبية كما يظهر في شكل (٣٠) . ويعطي الشكل الناتج النقط المئوية المختلفة للتوزيع . فمثلاً يلاحظ ان ٥٠٪ من الطلبة يقضون أقل من ٢١ ساعة أسبوعياً في ممارسة هواياتهم وان ٧٥٪ من الطلبة يقضون أقل من ٢٤ ساعة وان ٩٥٪ من الطلبة يقضون أقل من ٣١ ساعة وهكذا .

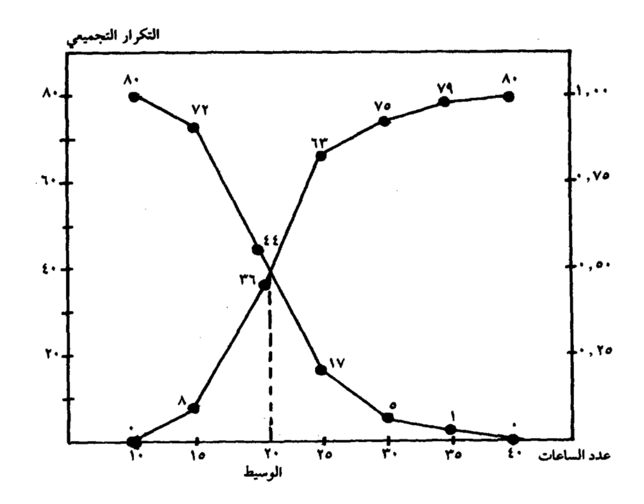
يمكن ان يرسم المضلع التجميعي الهابط بنفس الأسلوب . يعطي شكل (٣١) المضلع التجميعي الهابط المناظر لجدول (١٩) صفحة (٩٠) . ويلاحظ ان هذا الشكل قد رسم بطريقة تسمح بتمثيل التكرارات التجميعية المطلقة والتكرارات التجميعية النسبية في نفس الوقت ، إذ أن المحور الرأسي في الناحية اليسرى يعطي مقياس الرسم للتكرارات المطلقة على حين ان المحور الرأسي في الناحية اليمنى يعطي مقياس الرسم للتكرارات النسبية .



شكل (٣٠): المضلع التجميعي النسبي الصاعد لعينة من ٨٠ طالباً حسب عدد الساعات التي يقضونها في ممارسة هواياتهم .



شكل (٣١) التوزيع التكراري الهابط لعينة من ٨٠ طالباً حسب عدد الساعات التي يقضونها في ممارسة هواياتهم .



شكل (٣٢): المضلع التكراري الصاعد والمضلع التكراري الهابط لعينة من ٨٠ طالباً حسب عدد الساعات التي يقضونها في ممارسة هواياتهم

ويلاحظ من الشكل ان ٥٠٪ من الطلبة يقضون ٢١ ساعة على الأقل اسبوعياً في ممارسة هواياتهم وأن ٧٥٪ من الطلبة يقضون ١٨ ساعة على الأقل وهكذا .

يلاحظ عند رسم المضلع التكراري الصاعد والمضلع التكراري الهابط معاً في نفس الشكل ، ان مجموع التكرارين الصاعد والهابط المناظرين لقيمة ما من قيم المتغير يساوي دائماً المجموع الكلي للتكرارات (أو الواحد الصحيح في حالة رسم التكرارات النسبية) . وتمثل القيمة المناظرة لنقطة تقاطع المضلعين أحد المقاييس الاحصائية الهامة وهو الوسيط ، الذي سيناقش بالتفصيل فيما بعد . انظر شكل (٣٢) .

وتجدر الاشارة الى ان أسلوب رسم المضلعات التكرارية التجميعية لا يتأثر بعدم تساوي فئات التوزيع ، لأن ذلك يؤخذ في الاعتبار تلقائياً عند تجميع التكرارات . ولذلك لا يجب تعديل التكرارات عند رسم المضلعات التجميعية للتوزيعات ذات الفئات غير متساوية الطول .

وتجدر الاشارة الى ان استخدام الاضلاع للتوصيل بين النقط المختلفة في المضلع التجميعي يتضمن افتراض ان المشاهدات التي تقع داخل كل فئة تكون موزعة توزيعاً منتظماً داخل هذه الفئة . ولما كان هذا الافتراض تقريبي فان المضلع الناتج يمثل تقريباً للوضع الفعلي ، وغالباً ما يكون هذا التقريب كاف في معظم التطبيقات العملية .

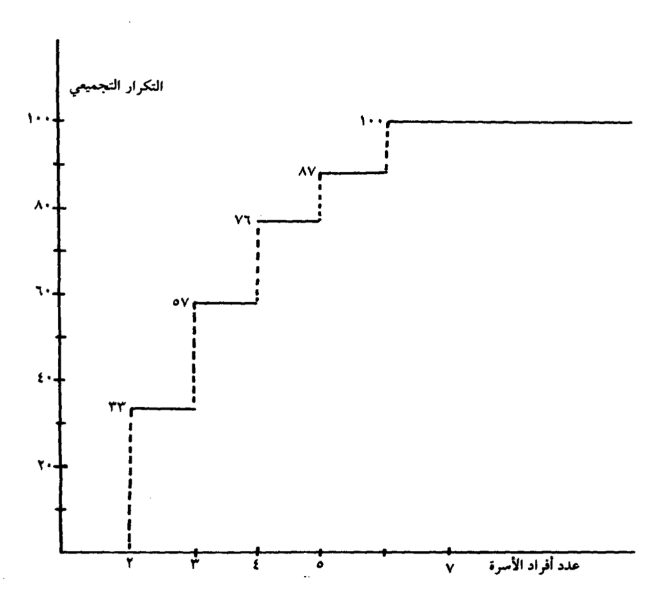
ويجب الاشارة الى بعض المواقف التي لا يمكن فيها توصيل النقط المختلفة في المضلع التجميعي بخطوط مستقيمة . وينشأ ذلك عند رسم هذه المضلعات لبيانات متقطعة . فمثلاً اذا كان توزيع عدد أفراد الأسرة في عينة من ١٠٠ أسرة يأخذ الشكل التالي :

عدد افراد الأسرة	٢	٣	٤	٥	٦	المجموع
عدد الأسرة	٣٣	٢٤	١٩	١١	١٣	١٠٠

فإن التوزيع التكراري التجميعي الصاعد يكون :

عدد افراد الأسرة	التكرار التجميعي
أقل من ٢	صفر
أقل من ٣	٣٣
أقل من ٤	٥٧
أقل من ٥	٧٦
أقل من ٦	٨٧
أقل من حد أعلى	١٠٠

ويأخذ المنحنى التجميعي المناظر في هذه الحالة شكل درجي كما يظهر في شكل (٣٣) . ويلاحظ ان عـدد الأفراد يجب ان يكـون عدداً صحيحـاً ، وبالتالي نجد ان المضلع يأخذ شكل درجة سلم عند كل قيمة من القيم .



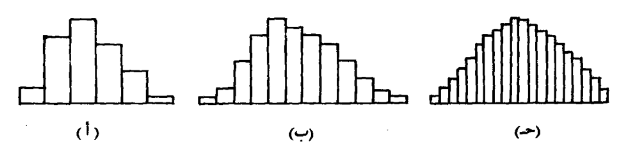
شكل (٣٣) : المضلع التجميعي الصاعد لعدد أفراد الأسرة في عينة من ١٠٠ أسرة

١٠ ـ المنحنى التكراري

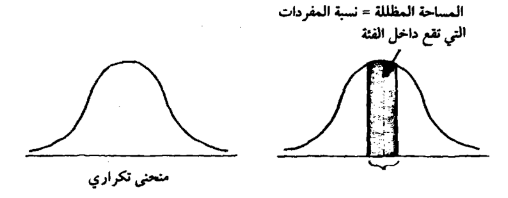
يتحدد شكل المدرج التكراري جزئياً في حالة المتغيرات المتصلة بكيفية اختيار فئات التوزيع المناظر . ويعتمد ذلك الى حد ما على حجم البيانات المستخدمة في انشاء التوزيع . ذلك انه من المنطقي استخدام عدد قليل من الفئات ذات طول واسع نسبياً ، اذا كانت البيانات تمثل عينة صغيرة من المجتمع . ويزيد عدد هذه الفئات ويقل طول كل منها كلما كبر حجم العينة ، وذلك في ضوء المعلومات الاضافية عن شكل توزيع مفردات المجتمع التي توفرها مثل هذه العينة . يعطي شكل (٣٤) أمثلة لمدرجات تكرارية نسبية تناظر عينات ذات احجام مختلفة من نفس المجتمع . ويلاحظ ان المساحة الكلية لكل مدرج في هذا الشكل تساوي الواحد . ويلاحظ ايضاً ان المدرج التكراري المناظر للعينة ذات الحجم الكبير يبدو اكثر نعومة وتمهيداً من المدرجات الأخرى . ويستنتج من ذلك انه كلما كبر حجم العينة المستخدمة فان عدد الفئات يصبح كبيراً وتكون اعمدة المدرج ذات قواعد صغيرة ، بحيث انه في النهاية يقترب شكل المدرج التكراري من شكل منحنى ممهد . ويسمى المنحنى الناتج بالمنحنى التكراري للمجتمع محل الدراسة . ويوضح شكل (٣٥) مثالاً لمنحنى تكراري . ويلاحظ انه طالما ان المنحنى التكراري يعتبر حالة خاصة من المدرج التكراري فان المساحة الكلية تحت المنحنى تساوي الواحد كما ان نسبة قيم المجتمع التي تقع داخل فئة ما تساوي المساحة تحت المنحنى المناظرة لتلك الفئة . وتمثل دراسة الأشكال والخصائص المختلفة للمنحنيات التكرارية جزءاً هاماً في علم الاحصاء . .

١١ ـ الشكل العام للمنحنى التكراري

يعتبر الشكل العام للمنحنى التكراري (أو المدرج التكراري) أحد السمات الأساسية التي يجب ملاحظتها عند دراسة نمط الاختلاف في التوزيع . ويمكن ان يوصف شكل المنحنى من خلال عدة خصائص أهمها :



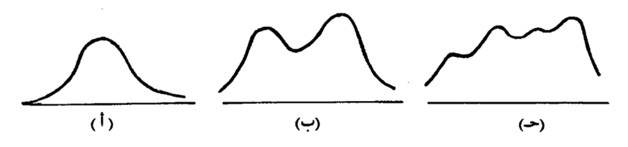
شكل (٣٤) : أمثلة لمدرجات تكرارية نسبية تناظر عينات ذات أحجام مختلفة من نفس المجتمع :
(أ) مدرج تكراري مناظر لعينة صغيرة الحجم
(ب) مدرج تكراري مناظر لعينة متوسطة الحجم
(جـ) مدرج تكراري مناظر لعينة كبيرة الحجم



شكل (٣٥): مثالاً لمنحنى تكراري

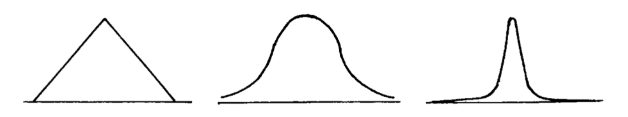
أ ـ **عدد القمم في المنحنى** . قد يكون المنحنى التكراري (أو المدرج التكراري) وحيد القمة Unimodal وقد يكون ثنائي القمة bimodal وقد يكون متعدد القمم multimodal . انظر شكل (٣٦) . وتجدر الاشارة الى ان المنحنى التكراري وحيد القمة هو المنحنى الأكثر شيوعاً في التطبيقات الاحصائية المختلفة . كذلك ، قد يظهر المنحنى ثنائي القمة في بعض الأحيان ، وغالباً ما تمثل البيانات في هذه الحالة مشاهدات عن مجموعتين غير متجانستين من المفردات . مثال ذلك المنحنى التكراري لأطوال الأشخاص البالغين في المجتمع ، حيث يمكن أن يكون للمنحنى قمة عند ١٦٠سم مثلاً تناظر القيمة الاكثر تكراراً للاناث ، وقمة اخرى عند ١٧٠سم مثلاً تناظر الطول

الأكثر تكراراً للذكور . اما المنحنى التكراري الذي تزيد عدد قممه عن قمتين ، فنادراً ما ينشأ عند وصف الظواهر الاحصائية المختلفة .



شكل (٣٦) منحنيات تكرارية مختلفة :
(أ) منحنى وحيد القمة
(ب) منحنى ثنائي القمة
(حـ) منحنى متعدد القمم

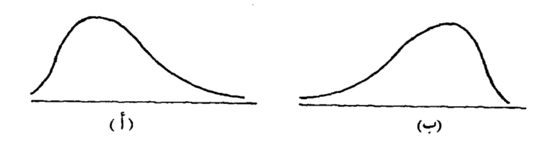
ب ـ **التماثل والالتواء** . يكون المنحنى التكراري (أو المدرج التكراري) متماثلاً Symmetric إذا كان هناك خطاً عمودياً ، يسمى خط التماثل ، يقسم المنحنى الى نصفين متشابهين تماماً . ويعطي شكل (٣٧) أمثلة لمنحنيات تكرارية متماثلة .



شكل (٣٧) : أمثلة لمنحنيات تكرارية متماثلة

اذا لم يكن المنحنى التكراري متماثلاً فإنه يكون ملتو Skewed . فإذا كان الذيل الأيمن للتوزيع يمتد بشكل أطول من امتداد الذيل الأيسر فإن المنحنى يكون ملتو لليمين (أو ملتو في الاتجاه الموجب) Positively Skewed . مثال ذلك المنحنى التكراري لتوزيع الدخل السنوي للأسر في احد المجتمعات ، حيث قد تتركز الغالبية العظمى من المشاهدات حول مركز

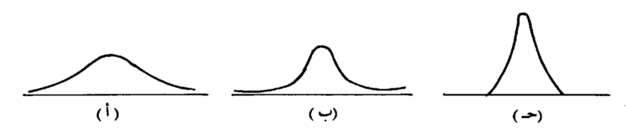
التوزيع مع وجود أعـداد قليلة من الأسر ذات دخـول عاليـة تؤدي الى سحب الذيل الأيمن للمنحنى . كذلك اذا امتد الذيل الأيسر للتوزيع بشكل أطول من امتداد الذيل الأيمن كان المنحنى ملتو لليسار (أو ملتـو في الاتجاه السالب) Negatively skewed . ومثـال ذلك المنحنى التكـراري لتوزيـع الدرجـات في امتحان ما اذا كان الامتحان سهلاً اذ قد تتركز معـظم الدرجـات قرب النهـاية العظمى للدرجات مع وجود عدد قليل نسبياً من الدرجات المنخفضة تؤدي الى امتداد الذيل الأيسر للمنحنى . انـظر شكل (٣٨) . وتجـدر الاشارة الى ان المنحنيات الملتوية لليمين اكثر انتشاراً في التطبيقات العملية من تلك الملتوية لليسار.



شكل (٣٨) : أمثلة لمنحنيات تكرارية ملتوية :
(أ) منحنى ملتو لليمين
(ب) منحنى ملتو لليسار

ولعل اكثر اشكال المنحنيات انتشاراً في التطبيقات الاحصائية المختلفة هو شكل المنحنى الطبيعي . ذلك ان هذا المنحنى يعد ملائماً لتمثيل توزيع العديد من الظواهر الاحصائية مثل طول الشخص ، مستوى الذكـاء ، مقياس ضغط الدم ، . . . الخ . هذا بالاضافة الى الأهميـة القصوى لهـذا المنحنى كأساس لكثيـر من عمليات الاستنتـاج الاحصائي . ويشبـه المنحنى الطبيعي المنحنى الأوسط في شكـل (٣٧) ، حيث يأخـذ شكلاً نـاقوسيـاً متماثـلاً ، ويكـون معدل التنـاقص على جـانبي النـاقـوس وفق قـانـون ريـاضي محـدد ومعروف .

وتجدر الاشارة الى ان اي منحنى نـاقوسي لا يكـون بالضـرورة منحنى طبيعياً . فمثلاً يلاحظ ان المنحنى في شكل ٣٩ (أ) هو منحنى طبيعي بينما المنحنى في شكل ٣٩ (ب) يختلف عن المنحنى الطبيعي في تناقص طرفي الناقوس بشكل أبطأ . كذلك فإن المنحنى في شكل ٣٩ (جـ) يتناقص طرفيه بمعدل أسرع من معدل تناقص طرفي المنحنى الطبيعي .



شكل (٣٩) أمثلة لمنحنيات ناقوسية :
(أ) منحنى طبيعي
(ب) منحنى ناقوسي يتناقص طرفيه بمعدل أبطأ من التوزيع الطبيعي
(جـ) منحنى ناقوسي يتناقص طرفيه بمعدل أسرع من التوزيع الطبيعي

ويعتمـد العمل الاحصـائي في معـظم الحـالات على مـدرج تكـراري لبيانات عينة للتعرف على سمات الشكل العام للمنحنى التكراري في المجتمع المنـاظر . ويتـطلب ذلك استخـدام أساليب الاستنتـاج الاحصائي التي تهتم بدراسة مدى التشابه بين المدرج التكراري المشاهد في العينة وبين المنحنى التكراري المناظر . كما تهتم بوصف الاختلافات المتوقعة في شكل المدرج التكراري من عينة لأخرى في نفس المجتمع . ويجب على القارىء أن يتذكر دائماً وجود أخطاء المعاينة التي تنعكس في اختلاف بيانات العينة عن بيانات المجتمـع وأن دراسـة وتحليـل هـذه الأخـطاء يمثـل جزءاً هـامـاً من العمـل الاحصائي .

١٢ - تحـويـل المتغيـرات الإحصـائيـة لتسهيـل وصف نمط الاختلاف

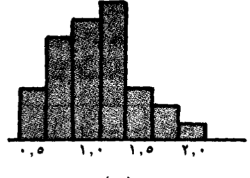
يهـدف التحليل الاحصـائي أساسـاً إلى استخدام طـرق سهلة وفعـالـة لوصف وتلخيص نمط الاختلاف في البيانات . اذا اتضح عند رسم المـدرج التكراري أن التوزيع الناتج معقد ويصعب تفسيره أو استخدامه ، فإنـه يمكن استخدام تحويله للمتغير محل الدراسة تؤدي إلى تبسيط هذا التوزيع . ويقصد بالتحويلة اجـراء عملية ريـاضية محـددة (مثل أخـذ الجذر التـربيعي أو أخذ اللوغـاريتم أو حساب المقلوب) على كـل مشـاهـدة من المشـاهـدات ، ثم تستخدم القيم الناتجة كأساس لاجراء الدراسة بدلاً من القيم الأصلية . وتختار التحويلة المستخدمة بهدف جعل توزيع القيم الناتجة اكثر تمـاثلاً من تـوزيع المشاهدات الأصلية ، وتكون التحويلة اكثر ملاءمة كلما اقترب التوزيع الناتج من شكل المنحنى الطبيعي . كمثال على ذلك ، يعطي جدول (٦) بيانات عن كمية الأمطار التي هطلت سنوياً على إحدى المناطق الصحراوية خلال الثلاثين عاماً الماضية (مقاسة بالسنتمتر) ، بالاضافة الى الجذر التـربيعي لكل كميـة من هذه الكميات .

ويظهر شكل (٤٠) المدرج التكراري المناظـر للقيم الأصلية والمـدرج التكراري المناظر للجذور التربيعية ، حيث يلاحظ أن توزيع القيم الأصلية ملتو لليمين بشكل واضح وأن توزيع الجذور التربيعية اكثر تماثلاً . كـذلك يمكن بسهولة التعرف على مركز ونمط الاختلاف لهذا التوزيع الأخير .

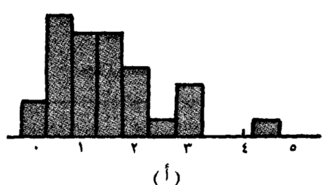
جدول (٦)

كمية الأمطار السنوية والجذر التربيعي لكمية الأمطار خلال الثلاثين عاماً الماضية في إحدى المناطق الصحراوية

السنة	كمية الأمطار	$\sqrt{\text{كمية الأمطار}}$	السنة	كمية الأمطار	$\sqrt{\text{كمية الأمطار}}$
١	٠,٧٧	٠,٨٨	١٦	١,٢٠	١,١٠
٢	١,٧٤	١,٣٢	١٧	٠,٤٧	٠,٦٩
٣	٠,٨١	٠,٩٠	١٨	١,٤٣	١,٢٠
٤	١,٢٠	١,١٠	١٩	٣,٣٧	١,٨٤
٥	١,٩٥	١,٤٠	٢٠	٢,٢٠	١,٤٨
٦	٣,٠٠	١,٧٣	٢١	٢,٨١	١,٦٨
٧	٣,٠٩	١,٧٦	٢٢	١,٨٧	١,٣٧
٨	١,٥١	١,٢٣	٢٣	١,١٨	١,٠٩
٩	٢,١٠	١,٤٥	٢٤	١,٣٥	١,١٦
١٠	٠,٥٢	٠,٧٢	٢٥	٤,٧٥	٢,١٨
١١	١,٦٢	١,٢٧	٢٦	٢,٤٨	١,٥٧
١٢	١,٣١	١,١٤	٢٧	٠,٩٦	٠,٩٨
١٣	٠,٣٢	٠,٥٧	٢٨	١,٨٩	١,٣٧
١٤	٠,٥٩	٠,٧٧	٢٩	٠,٩٠	٠,٩٥
١٥	٠,٨١	٠,٩٠	٣٠	٢,٠٥	١,٤٣



(ب)

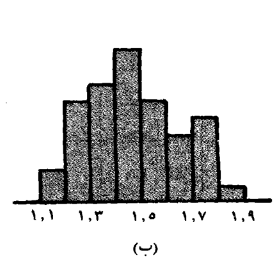


(أ)

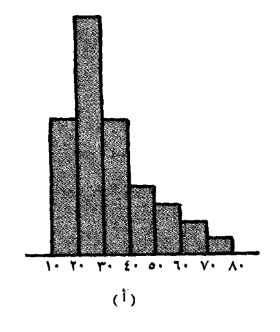
شكل (٤٠): المدرج التكـراري المناظـر لتوزيـع كمية الأمـطار التي هطلت على احـدى المناطق الصحراوية خلال الثلاثين عاماً الماضية :

(أ) البيانات الأصلية

(ب) الجذور التربيعية للبيانات الأصلية



(ب)



(أ)

شكل (٤١): المدرج التكراري لبيانات الزمن الذي يستغرق كل من ٤٠ شخص في العدو لمسافة ٥٠ متراً :

(أ) البيانات الأصلية

(ب) لوغاريتمات البيانات الأصلية .

يعطي جدول (٧) مثالاً آخر لاستخدام التحويلات لوصف المتغيرات الاحصائية ، حيث تمثل البيانات الزمن الذي يستغرقه كل من ٤٠ شخصاً في العدو لمسافة ٥٠ متراً (مقاساً بالثوان). استخدمت تحويلة اللوغاريتم على هذه البيانات ، ويظهر الجدول القيم الأصلية ولوغاريتماتها .

جدول (٧)

الزمن الذي يستغرقه الشخص في العدو مسافة ٥٠ متراً ولوغارتيمات هذه القيم

الشخص	الزمن	لوالزمن	الشخص	الزمن	لوالزمن	الشخص	الزمن	لوالزمن
١	٢٨,١	١,٤٥	١٥	٦٠,١	١,٧٨	٢٩	٢١,٠	١,٣٢
٢	٣١,٢	١,٤٩	١٦	٢٣,٧	١,٣٧	٣٠	٢٢,٣	١,٣٥
٣	١٣,٧	١,١٤	١٧	١٨,٦	١,٢٧	٣١	١٥,٥	١,١٩
٤	٤٦,٠	١,٦٦	١٨	٢١,٤	١,٣٣	٣٢	٣٦,٣	١,٥٦
٥	٢٥,٨	١,٤١	١٩	٢٦,٦	١,٤٢	٣٣	١٩,١	١,٢٨
٦	١٦,٨	١,٢٣	٢٠	٣٢,٠	١,٥١	٣٤	٣٨,٤	١,٥٨
٧	٣٤,٨	١,٥٤	٢١	٤٣,٥	١,٦٤	٣٥	٧٢,٨	١,٨٦
٨	٦٢,٣	١,٧٩	٢٢	١٧,٤	١,٢٤	٣٦	٤٨,٩	١,٦٩
٩	٢٨,٠	١,٤٥	٢٣	٣٨,٨	١,٥٩	٣٧	٢١,٤	١,٣٣
١٠	١٧,٩	١,٢٥	٢٤	٣٠,٦	١,٤٩	٣٨	٢٠,٧	١,٣٢
١١	١٩,٥	١,٢٩	٢٥	٥٥,٦	١,٥٧	٣٩	٥٧,٣	١,٧٦
١٢	٢١,١	١,٣٢	٢٦	٢٥,٥	١,٤١	٤٠	٤٠,٩	١,٦١
١٣	٣١,٩	١,٥٠	٢٧	٥٢,١	١,٧٢			
١٤	٢٨,٩	١,٤٦	٢٨	٢٦,٢	١,٤٢			

ويوضح شكل (٤١) المدرج التكراري المناظر لقيم الزمن الأصلية والمدرج التكراري المناظر للوغاريتمات ، حيث يلاحظ أن القيم الأصلية ملتوية في اتجاه اليمين بشكل واضح على حين أن توزيع اللوغاريتمات يأخذ شكلاً متماثلاً يقترب من التوزيع الطبيعي .

ويتطلب التوصل إلى التحويلة المناسبة اتباع أسلوب التجربة والخطأ في كثير من الأحيان ، وذلك برسم عدد من التحويلات ثم اختيار أفضلها . وقد يعتمد في هذا الصدد على الحاسبات الآلية التي تمكن من إتمام عملية الاختيار بسرعة وكفاءة .

تمرينات

١ - ارسم التوزيع التكراري النسبي في التمرين رقم (١) صفحة (١٠٣) في شكل بياني مناسب .

٢ - إرسم التوزيع التكراري في التمرين رقم (٢) صفحة (١٠٣) في شكل بياني مناسب .

٣ - ارسم شكلاً بيانياً مناسباً لتمثيل الجدول التكراري المزدوج في التمرين رقم (٣) صفحة (١٠٤) .

٤ - ارسم شكلاً بيانياً مناسباً لتمثيل الجدول التكراري المزدوج في التمرين رقم (٢٠) صفحة (١١٢).

٥ - ارسم شكلاً بيانياً مناسباً لتمثيل الجدول الذي يظهر في التمرين رقم (٢٥) صفحة (١١٥).

٦ - فيما يلي عدد الوفيات في دولة ما خلال الشهور المختلفة لعام ١٩٨٥ :

الشهر	يناير	فبراير	مارس	أبريل	مايو	يونيه
عدد الوفيات بالألف	١٦٧	١٥١	١٦٥	١٥٩	١٥٦	١٤٩
الشهر	يوليه	أغسطس	سبتمبر	اكتوبر	نوفمبر	ديسمبر
عدد الوفيات بالألف	١٦٠	١٤٥	١٤١	١٥٥	١٥١	١٦٤

(أ) ارسم هذه البيانات في شكل مناسب .

(ب) هل تعتقد أن مستوى الوفاة في هذه الدولة يختلف من شهر لآخر خلال السنة ؟ اشرح سبب إجابتك .

٧ - فيما يلي توزيع السكان حسب العمر في كل إمارة من امارات الدولة كما ورد في تعداد السكان لعام ١٩٨٠ .

الامارة	أبو ظبي	دبي	الشارقة	عجمان	أم القيوين	رأس الخيمة	الفجيرة
عدد الذكور بالألف	٣٣٢	١٨٨	١٠٣	٢٢	٨	٤٧	٢١
عدد الاناث بالألف	١٢٠	٨٩	٥٧	١٤	٤	٢٧	١١

أ ـ مثل هذه البيانات بشكل بياني مناسب .

ب ـ ما هي الأسباب المحتملة في رأيك لعدم التوازن الملاحظ بين اعداد الذكور وأعداد الاناث ؟

٨ ـ فيما يلي توزيع السكان حسب العمر والنوع في الدولة كما ورد في تعداد ١٩٨٠ . (الأعداد بالآلاف) .

فئات العمر	عدد الذكور	عدد الاناث
صفر ـ ٩	١٢١	١١٤
١٠ ـ ١٩	٦٥	٥٤
٢٠ ـ ٢٩	٢٣٥	٧٣
٣٠ ـ ٣٩	١٨٧	٤٣
٤٠ ـ ٤٩	٧٦	١٩
٥٠ ـ ٥٩	٢٣	١٠
٦٠ ـ ٦٩	٨	٥
٧٠ ـ ٧٩	٣	٢
٨٠ +	١	١

أ ـ مثل هذه البيانات بشكل بياني مناسب

ب ـ احسب نسبة الذكور إلى الإناث في كل فئة عمرية ، ومثل الناتج بشكل بياني مناسب ، وعلق على الأسباب المحتملة للنمط المشاهد في هذه البيانات .

٩ ـ فيما يلي توزيع السكان (١٠ سنوات فأكثر) في الدولة حسب الحالة التعليمية في عام ١٩٨٠ . (الأعداد بالآلاف) .

الحالة التعليمية	أمي	يقرأ ويكتب	ابتدائية	اعدادية	ثانوية	دون الجامعية	جامعية
عدد السكان بالألف	٢٥٦	١٨٢	٩٨	٨٢	١١٠	٢١	٥٢

والمطلوب تمثيل هذه البيانات بشكل بياني مناسب .

١٠ - فيما يلي توزيع السكان البالغين في الدولة حسب الحالة الزواجية والنوع في عام ١٩٨٠ .(الأعداد بالألف) .

الحالة الزواجية	ذكور	اناث
لم يسبق له الزواج	١٧٠	٢٥
متـزوج	٣٩٢	١٣٩
مـطلق	٢	٣
أرمـــل	٢	١٢

والمطلوب تمثيل هذا الجدول بشكل بياني مناسب .

١١ - فيما يلي توزيع عدد المواليد حسب النوع في شهور السنة المختلفة عام ١٩٨٣ والمطلوب عرض هذا الجدول بشكل بياني مناسب :

الشهر	١	٢	٣	٤	٥	٦
عدد المواليد الذكور	١٩٠٨	١٧٠٣	١٧٧٨	١٧٩٠	١٧٠١	١٧٦٦
عدد المواليد الاناث	١٨٦٨	١٥٦٥	١٦٩٩	١٦٨٩	١٦٣٣	١٧٠٨
الشهر	٧	٨	٩	١٠	١١	١٢
عدد المواليد الذكور	١٨٨٥	١٩١٤	١٨٤٥	٢٠٤٩	١٨٨٠	١٩٤٣
عدد المواليد الاناث	١٧٤٨	١٨٩٠	١٨٣٨	١٩١١	١٨٩٢	١٨١٦

١٢ - فيما يلي بيان بالمساحات التي تم تشجيرها وعدد الأشجار المغروسة في منطقة العين (لاحظ أن الأعداد تراكمية توضح المساحات وعدد الأشجار المغروسة فعلاً) وذلك في السنوات ١٩٧٨ - ١٩٨٤ . والمطلوب عرض هذه البيانات في شكل مناسب .

السنة	المساحات بآلاف الهكتار	عدد الأشجار بالآلاف
١٩٧٨	١٠,٦	٢١٢٣
١٩٧٩	١٠,٨	٢١٦٩
١٩٨٠	١٠,٤	٢٢٨١
١٩٨١	١١,٧	٢٣٤٧
١٩٨٢	١٢,٤	٢٤٧٧
١٩٨٣	١٣,٤	٢٦٨٩
١٩٨٤	١٣,٨	٢٧٨٩

١٣ ـ فيما يلي بيان بحوادث المرور المسجلة في إمارة أبو ظبي في عامي ٨٣، ١٩٨٤ حسب نوع الحادث .

السنة	اصطدام	تدهور	دهس	الجملة
١٩٨٣	٥٧٣٢	١٧٦٤	٤٩٣	٧٩٨٩
١٩٨٤	٤٣٣٩	١٤٢٥	٣١٧	٦١٨١

والمطلوب عرض هذا الجدول في شكل مناسب .

١٤ ـ فيما يلي بيان بعدد المنشآت غير الحكومية وعدد العاملين بها في امارات الدولة خلال عام ١٩٨٠ ، والمطلوب عرض هذه البيانات بشكل مناسب .

الامارة	عدد المنشآت	عدد المشتغلون
ابو ظبي	١٣٢٠٧	١٥٦٥٠٥
دبي	١٠٩٥٢	١٢٩٧٢٠
الشارقة	٦٦٢٥	٥٤٦١٤
عجمان	١٦٧١	٨١١٦
أم القيوين	٣٣٨	٢٣١٩
رأس الخيمة	٢٩٤٤	١٦١١٩
الفجيرة	١٠٦٠	٧٦٢٥

١٥ ـ (أ) ارسم المدرج التكراري المناظر للجدول التكراري في التمرين رقم (٥) صفحة (١٠٥) ، وعلق على شكل المدرج الناتج .

(ب) ارسم المضلع التكـراري التجميعي الصـاعـد المنـاظـر لهـذه البيانات .

١٦ ـ (أ) ارسم المدرج التكراري الـنسبي المنـاظر للجـدول التكراري في التمرين رقم (٧) صفحة (١٠٦) وعلق على الشكل الناتج .

(ب) ارسم المضلع التكراري التجميعي الهابط المناظر لهذه البيانات .

١٧ ـ يعطي الجدول الآتي التوزيع التكراري لعدد الجرائم لمجموعة من ٢٨٣ سجيناً :

عدد الجرائم	١	٢	٣	٤	٥	٦	٧	٨	٩	١٠	١١	المجموع
عدد السجناء	١٦	٢٧	٣٧	٤٦	٣٦	٤٠	٣١	٢٧	١٣	٨	٢	٢٨٣

(أ) ارسم المدرج التكراري النسبي المناظر .

(ب) ارسم المضلع التجميعي النسبي الصاعد المناظر لهذه البيانات .

١٨ ـ يعـطي الجدول الآتي التـوزيع التكـراري لعدد الأبحـاث التي قـام بهـا خريجو إحدى الجامعات منذ التخرج .

عدد الأبحاث	١	٢	٣	٤	٥	٦	٧	٨
عدد الخريجين	٧٨٤	٢٠٤	١٢٧	٥٠	٣٣	٢٨	١٩	١٩

٩	١٠	١١	١٢	١٣	١٤	١٥	١٦	١٧	المجموع
٦	٧	٦	٧	٤	٤	٥	٣	٣	١٢٢٩

ارسم المدرج التكراري الـنسبي المناظر ، وعلق على شكـل المدرج الناتج .

١٩ ـ استخدم بيانات التمرين رقم (١٥) صفحة (١٠٩) للاجـابة عن الاسئلة الآتية :

(أ) ارسم المدرج التكراري النسبي المناظر .

(ب) ارسم المضلع التكراري النسبي المناظر .

(حـ) ارسم المضلع التجميعي النسبي الصاعد واستخدمه لإيجاد الدرجة التي يقل عنها ٥٠٪ من الدرجات .

(ء) ارسم المضلع التجميعي النسبي الهابط واستخدمه لايجاد الحد الأدنى للدرجات التي حصل عليها أفضل ١٠٪ من الطلبة .

٢٠ ـ (أ) ارسم المضلع التكراري النسبي المناظر لبيانات التمرين رقم (١٦) صفحة (١١٠) .

(ب) ارسم المضلع التجميعي النسبي الصاعد المناظر لهذه البيانات ، واستخدمه لحساب نسبة الأجهزة التي تصاب بعطل خلال سنة واحدة من تركيبها .

٢١ ـ استخدم بيانات التمرين رقم (١٧) صفحة (١١١) للإجابة عن الاسئلة الآتية :

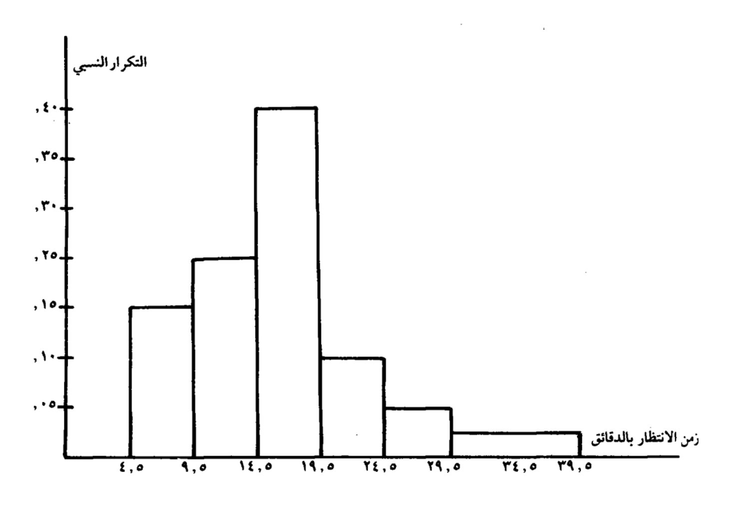
(أ) ارسم المضلع التكراري النسبي المناظر .

(ب) ارسم كلا من المضلع التجميعي النسبي الصاعد والمضلع التجميعي النسبي الهابط وحدد العمر المناظر لنقطة تقاطعهما . ما هو تفسيرك لمعنى هذا العمر ؟

(حـ) اوجد العمر الذي يمثل الحد الأدنى لأعمار الربع الأكبر سناً من الموظفين .

(ء) اوجد العمر الذي يمثل الحد الأعلى لأعمار الخمس الأقل سناً من الموظفين .

٢٢ ـ يعطي شكل (٤٢) المدرج التكراري النسبي المناظر للتوزيع التكراري للزمن الذي ينتظره الشخص في محطة القطارات حتى وصول قطاره وذلك لعينة من ٣٠٠ شخص . كون الجدول التكراري المناظر لهذا المدرج .



شكل (٤٢): التوزيع النسبي لأزمنة انتظار ٢٠٠ شخص في محطة القطارات

٢٣ ـ يعطي الجدول التالي التوزيع التكراري لعدد الكيلومترات التي تقطعها الحافلة قبل حدوث أول عطل فني لها وذلك لمجموعة من ١٩١ حافلة تعمل على خطوط المواصلات العامة في الدولة .

فئات المسافة بآلاف الكيلومترات	صفر ـ ١٩	٢٠ ـ ٣٩	٤٠ ـ ٥٩	٦٠ ـ ٧٩	٨٠ ـ ٩٩	١٠٠ ـ ١١٩	١٢٠ ـ ١٣٩	١٤٠ ـ ١٥٩	١٦٠ ـ ١٧٩	١٨٠ ـ ١٩٩	المجموع
عدد الحافلات	٦	١١	١٦	٢٥	٣٤	٤٦	٣٣	١٦	٢	٢	١٩١

(أ) ارسم المدرج التكراري وعلق على الشكل الناتج .

(ب) اوجد نسبة الحافلات التي تسير لمسافة ٥٠ الف كيلومتراً على الأقل قبل حدوث أول عطل فني لها .

(حـ) اوجد نسبة الحافلات التي يحدث لها أول عـطل فني قبل أن تبلغ المسافة التي تقطعها كل منها ٩٠ الف كيلومتراً .

(ء) اوجد نسبة الحافلات التي يحدث لها أول عـطل فني بعد سيـرها لمسافة تتراوح بين ٧٠ألف كيلومتراً ، ١٥٠ ألف كيلومتر .

٢٤ ـ تتعلق البيانات التـالية بـدراسة سبقت الاشـارة إليها في التمـرين رقم ٣ صفحة (١٠٤) . فيما يلي التوزيع العمري للأشخاص في البيانات التي جمعت بعـد زيارة واحـدة والتوزيـع العمري لـلأشخاص في البيـانـات المتوافرة بعد عشر زيارات .

فئات العمر بالسنوات	التكرار النسبي بعد زيارة واحدة	التكرار النسبي بعد عشر زيارات
١٨ ـ ٢١	,٠٦٨	,٠٧١
٢٢ ـ ٢٩	,٢١٣	,٢٤٢
٣٠ ـ ٣٩	,١٨٧	,٢٣٠
٤٠ ـ ٤٩	,١١١	,١٣٠
٥٠ ـ ٥٩	,١٢٣	,١٣٣
٦٠ ـ ٧٩	,٢٩٧	,١٩٦

(أ) ارسم المدرج التكراري المناظر لكل توزيع .

(ب) ارسم المضلعـات التكراريـة المناظـرة في نفس الشكـل وقـارن بينهما .

٢٥ ـ يعطي الجدول التالي التوزيع التكراري للزمن الذي يقضيه الـطبيب في فحص كل مريض وذلك لكل من الطبيب أ والطبيب ب وذلك خلال يوم معين .

فئات الزمن بالدقائق	عـــدد المـــرضـــى	
	للطبيب أ	للطبيب ب
٢ - ٦	١٢	١
٧ - ١١	١٣	٣
١٢ - ١٦	٨	٤
١٧ - ٢٧	٦	٨
٢٨ أو أكثر	١	٤
المجموع	٤٠	٢٠

ارسم المضلعات التكرارية النسبية المناظرة وقارن بينها .

٢٦ - يراد إنشاء التوزيع التكراري للدرجات التي قد يحصل عليها ٧٠ طالباً في أحد الاختبارات باستخدام الفئات : ١٠٠ - ١١٩، ١٢٠ - ١٣٩، ١٤٠ - ١٥٩، ١٦٠ - ١٧٩، ١٨٠ - ١٩٩ . وزع الدرجات على الفئات المختلفة لتحقيق الشروط التالية :

(أ) يكون التوزيع الناتج متماثلاً .
(ب) يكون التوزيع الناتج ملتو لليمين .
(حـ) يكون التوزيع الناتج ملتو لليسار .
(ء) يكون التوزيع الناتج ذو قمتين .

٢٧ - يهدف هذا التمرين إلى توضيح الأشكال المختلفة للتوزيعات التكرارية .

فيما يلي خمس توزيعات تكرارية افتراضية :

الفئات	التوزيع الأول	التوزيع الثاني	التوزيع الثالث	التوزيع الرابع	التوزيع الخامس
صفر - ٩	٥	٤٠	٣٠	٥	٦
١٠ - ١٩	١٠	٢٥	١٠	٢٥	٥
٢٠ - ٢٩	٢٠	١٠	٨	٨	٦
٣٠ - ٣٩	٣٠	٨	٧	٧	٩
٤٠ - ٤٩	٢٠	٧	٧	٢٠	٩
٥٠ - ٥٩	١٠	٥	٨	٢٥	٢٣
٦٠ - ٦٩	٥	٥	٣٠	١٠	٤٢

ارسم المدرج التكراري المناظر لكل توزيع وعلق على كل شكل من حيث التماثل ، واتجاه الالتواء ، وعدد قمم التوزيع .

٢٨ - استخدم البيانات الخاصة بكمية الأمطار السنوية المعطاة كمثال في صفحة (١٦٣) ، واحسب الجذر التكعيبي لكل قيمة من هذه القيم .
ارسم المدرج التكراري النسبي المناظر لهذه الجذور وعلق على شكله .
هل تعتقد أن استخدام تحويلة الجذر التكعيبي تساعد في وصف البيانات الأصلية بطريقة أسهل ؟

٢٩ - في دراسة تتعلق بأنماط استهلاك أنابيب معجون الأسنان في مجتمع ما ، اختيرت عينة عشوائية من ٢٠٧١ أسرة وطلب من كل منها تسجيل عدد الأنابيب المشتراة خلال فترة الدراسة (التي امتدت لمدة أربع سنوات) . فيما يلي التوزيع التكراري لعدد أنابيب معجون الأسنان التي اشترتها هذه الأسر .

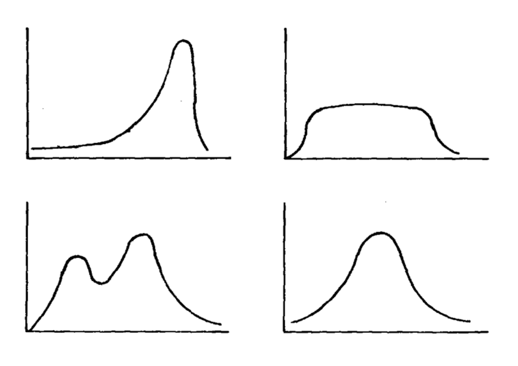
عدد الانابيب المشتراة	١٠ - ١٩	٢٠ - ٢٩	٣٠ - ٣٩	٤٠ - ٤٩	٥٠ - ٥٩	٦٠ - ٦٩	٧٠ - ٧٩	٩٠ - ١٠٩	١١٠ - ١٢٩	١٣٠ - ١٤٩	١٥٠ - ١٦٩	المجموع
عدد الأسر	٩٠٤	٥٠٠	٢٥٨	١٦٧	٩٤	٥٦	٤٦	٢٢	١٣	٩	٢	٢٠٧١

(أ) ارسم المدرج التكراري النسبي المناظر ، وعلق على الشكل العام لهذا المدرج .

(ب) كون التوزيع التكراري الذي ينشأ باستخدام تحويلة الجذر التربيعي لهذه البيانات .

(حـ) ارسم المدرج التكراري النسبي المناظر للجذور التربيعية . علق على مدى ملاءمة تحويلة الجذر التربيعي لهذه البيانات .

٣٠ ـ اكتب وصفاً موجزاً للشكل العام لكل من المنحنيات التكرارية الآتية التي تظهر في شكل (٤٣) .



شكل (٤٣) : منحنيات تكرارية مختلفة

٣١ ـ يعطي الجدول التالي التوزيع النسبي لأرقام الآحاد في الأعمار المسجلة للسكان البالغين في تعداد السكان لكل من البلد (أ) والبلد (ب) .

رقم الآحاد	صفر	١	٢	٣	٤
نسبة السكان في البلد أ	٠,١٦٨	٠,٦٧	,٠٩٤	٠,٠٨٦	,٠٨٨
نسبة السكان في البلد ب	,١٠٦	,٠٩٩	,١٠٠	,٠٩٦	,٠٩٨

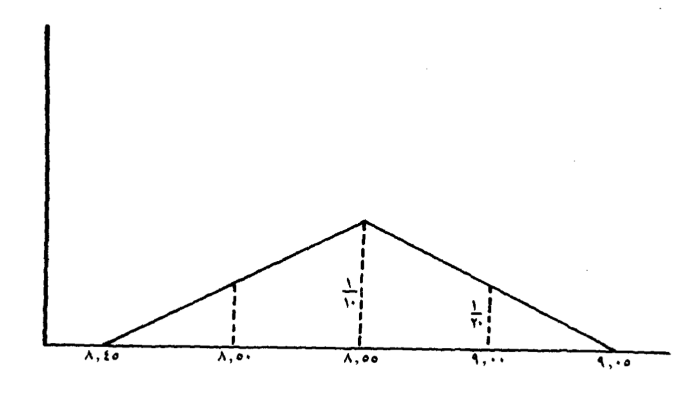
رقم الآحاد	٥	٦	٧	٨	٩	المجموع
نسبة السكان في البلد أ	,١٣٤	,٠٩٤	,٠٨٥	,١٠٢	,٠٨٢	١,٠٠٠
نسبة السكان في البلد ب	,١٠٠	,٠٩٩	,١٠٢	,١٠٠	,١٠١	١,٠٠٠

أ ـ ارسم المدرج التكراري لسكان البلد أ والمدرج التكراري لسكان البلد ب في شكل واحد .

ب ـ علق على الاختلافات المشاهدة بين سكان البلدين . ما هي في رأيك الأسباب التي تؤدي إلى وجود هذه الاختلافات .

حـ ـ يعتقد المشتغلون بالدراسات السكانية أن الأشخاص عند تسجيل أعمارهم يفضلون اعطاء أعمار تنتهي بالصفر أو الخمسة ، كما يفضلون الأعمار التي تنتهي بأرقام زوجية . هل تدل البيانات المعطاة على صحة هذا الاعتقاد ؟ اشرح سبب اجابتك .

٣٢ ـ يحاول الموظفون الوصول إلى مكاتبهم كل صباح قبل موعد بدء الدوام الرسمي وهو تمام الساعة التاسعة صباحاً . جمعت بيانات عن مواعيد الوصول لجميع الموظفين ووضعت هذه البيانات في توزيع تكراري نسبي يأخذ الشكل الآتي :



أ ـ ما هو المدى الزمني الذي يصل خلاله الموظفون إلى مكاتبهم ؟

ب ـ تأكد من أن المساحة الكلية تحت المنحنى التكراري النسبي تساوي الواحد .

جـ ـ علق على الشكل العام لهذا المنحنى التكراري .

ء ـ احسب نسبة الموظفين الذين يتأخرون في الوصول إلى عملهم كل صباح .



الباب الخامس

مدخل الى المقاييس الوصفية

١ ـ مقدمة

رأينا في الأبواب السابقة كيف يمكن وصف نمط الاختلاف في مجموعة بيانات من خلال تنظيم هذه البيانات ووضعها في توزيع تكراري . ويتم عرض التوزيع التكراري في شكل جدول أو رسم بياني ، وهي أشكال تخاطب الحس البصري للقارىء بهدف إعطائه معلومات عن السمات الأساسية للتوزيع . ويمكن القول بأن السمات الأساسية لمعظم التوزيعات التكرارية هي :

أ ـ وجود قيمة متوسطة تتركز حولها معظم المشاهدات . ويعبر عن ذلك بوجود نزعة في البيانات نحو المركز .

ب ـ وجود نمط لاختلافات المشاهدات فيما بينها . ويشمل ذلك المدى الذي تنتشر عليه القيم واسلوب تشتت هذه القيم داخل المدى .

حـ ـ الشكل العام للمنحنى التكراري المناظر ويدخل في ذلك شكل قمة المنحنى ، ودرجة وأسلوب التماثل في المنحنى ، واتجاه الالتواء إن وجد .

يمكن قياس كل من هذه السمات بمقياس عددي مناسب . ويترتب على ذلك إمكانية الحصول على وصف تقريبي للتوزيع التكراري من خلال عدد من المقاييس الوصفية التي يمكن تفسير معناها . ولا يعني ذلك بالطبع أن تستخدم هذه المقاييس كبديل كامل عن التوزيع التكراري وإنما المقصود هو إمكانية

الاعتماد على هذه المقاييس للحصول على معلومات مفيدة عن الخصائص الرئيسية للتوزيع .

وتجدر الإشارة إلى الأهمية العظمى لهذه المقاييس الوصفية في عملية التحليل الاحصائي للبيانات . ويرجع ذلك للأسباب الآتية :

أ - يمكن أن يتم حساب هذه المقاييس في كثير من الأحيان على نحو أسرع وأكثر سهولة من انشاء التوزيع التكراري .

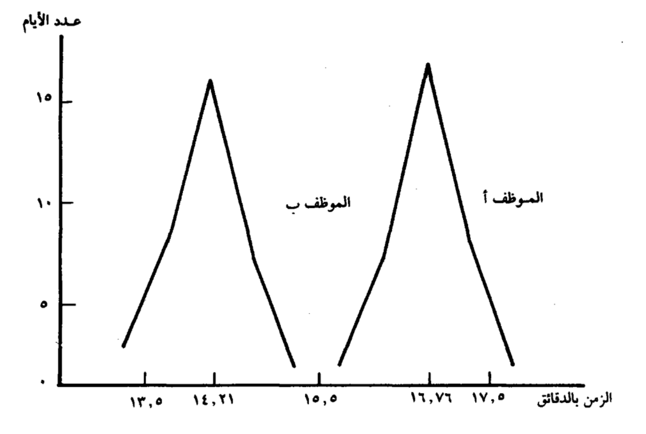
ب - قد تكون هذه المقاييس اكثر ملاءمة من التوزيع التكراري الكامل في المواقف التي يكون عدد المشاهدات فيها صغيراً .

حـ - أن وضع التوزيع التكراري في شكل جدول أو رسم بياني لا يسمح باستخدامه لأغراض التحليل الرياضي في المراحل التالية للعمل الاحصائي . ويكتسب ذلك أهمية خاصة في عمليات الاستنتاج الإحصائي ، حيث يتم اخضاع نتائج العينة لعمليات تحليلية بهدف استنتاج خصائص المجتمع ، ويكون من الضروري في هذه الحالة الاعتماد على المقاييس الوصفية التي تحسب من نتائج العينة .

نناقش فيما يلي ، بشكل عام ، الأنواع المختلفة للمقاييس الوصفية ، وهي مقاييس المركز (أو الموضع) ومقاييس التشتت ومقاييس الالتواء ومقاييس التفرطح . أما المناقشة التفصيلية التي تشمل كيفية حساب وتفسير هذه المقاييس فتترك للأبواب التالية .

٢ - مقاييس الموضع (أو المركز)

تستخدم المتوسطات بكثرة لوصف التوزيع التكراري . ولعل اكثر هذه المتوسطات شهرة هو الوسط الحسابي لمجموعة من القيم (الذي يحسب بجمع هذه القيم والقسمة على عددها) . يعطي شكل (١) مثالاً لاستخدام الوسط الحسابي في التعبير عن خصائص التوزيعات التكرارية ، حيث تظهر التوزيعات التكرارية للزمن الذي يستغرقه كل من الموظف أ والموظف ب في



شكل (١) : التوزيعات التكرارية للزمن الذي يستغرقه كل من الموظف أ والموظف ب في الذهاب إلى عملهم كل صباح .

الذهاب إلى عملهم كل صباح . ويلاحظ ما يلي .

أ - إذا نظرنا إلى توزيع أزمنة الموظف أ ، يلاحظ أن الوسط الحسابي لهذه الأزمنة هو ١٦,٧٦ دقيقة . ويلاحظ كذلك أن هذا الوسط يقع في مركز البيانات ، بحيث أن اختلافات معظم المشاهدات عن هذا الوسط تبدو ضئيلة . ويدل ذلك على أن الزمن الذي يستغرقه الموظف أ في الذهاب إلى عمله كل صباح يكون قريباً في معظم الأيام من المتوسط ١٦,٧٦ دقيقة ، وأن هناك نزعة في المشاهدات نحو التركز حول هذا المتوسط . ويقال في هذه الحالة أن المتوسط مقياس لموضع التوزيع أو مقياس للنزعة المركزية في التوزيع .

ب - يلاحظ أن الشكل العام لتوزيع أزمنة الموظف أ مشابه للشكل العام لتوزيع أزمنة الموظف ب ، وأن الوسط الحسابي مقياس جيد للموضع في كلا الحالتين . ويمكن تبعاً لذلك الاعتماد على الوسط الحسابي للمقارنة بين موضعي التوزيعين . فمثلاً نجد أن الوسط الحسابي لتوزيع

(أ) هو ١٦,٧٦ دقيقة بينما الوسط الحسابي لتوزيع (ب) هو ١٤,٢١ دقيقة مما يعني في هذه الحالة أن توزيع (ب) يقع دائماً إلى يسار توزيع (أ) وأن الموظف (ب) يتطلب زمناً أقل للوصول إلى عمله من الزمن الذي يتطلبه الموظف أ .

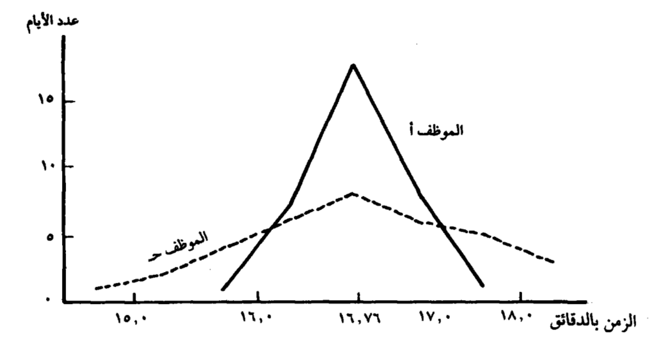
ويجب التأكيد على أن المتوسطات لا تكون دائماً مقاييس جيدة للموضع . ويتضح ذلك من شكل (٢) الذي يعطي المنحنى التكراري لتوزيع أزمنة الموظف (أ)مع المنحنى التكراري لتوزيع أزمنة موظف ثالث (حـ). يلاحظ أن الوسط الحسابي متساو للتوزيعين ويساوي ١٦,٧٦ دقيقة ، ولكن لا يمكن الادعاء بأن التوزيعين لهما نفس الموضع . ويرجع الانخفاض في كفاءة المتوسط كمقياس للموضع في هذه الحالة إلى اعتبارين متداخلين هما :

أ ـ أن هناك اختلافات واسعة في أزمنة الموظف (حـ) ، بحيث لا يوجد تركز واضح لهذه الأزمنة حول المتوسط . ويدل ذلك على عدم وجود نزعة مركزية واضحة في البيانات ، ولا يمكن تبعاً لذلك الاعتماد على الوسط وحده لوصف موضع هذا التوزيع .

(ب) أن هناك اختلافات أساسية في نمط التشتت بين أزمنة الموظف (أ) وأزمنة الموظف(حـ)، وبالتالي لا يمكن الاعتماد على المتوسط فقط لمقارنة موضعي التوزيعين .

ونخلص مما سبق إلى ما يلي عند وصف موضع التوزيع التكراري :

أ ـ يستخدم المتوسط كمقياس للموضع عندما يكون هناك نزعة مركزية واضحة في التوزيع ، أي عندما لا تختلف المشاهدات فيما بينها اختلافاً كبيراً . وعلى ذلك فإنه ينبغي عند استخدام المتوسطات أن تتوافر معلومات اضافية عن حجم الاختلافات بين المشاهدات ، حتى يمكن الحكم على جودة هذه المتوسطات كمقاييس للموضع . وتأتي هذه المعلومات إما بالنظر إلى خصائص التوزيع التكراري أو بحساب مقاييس



شكل (٢) : التوزيعات التكرارية للزمن الذي يستغرقه كل من الموظف أ والموظف جـ في الذهاب إلى عملهم كل صباح .

ملائمة لتشتت المشاهدات ، وستناقش هذه المقاييس فيما بعد .

ب ـ لا يستخدم المتوسط كأساس للمقارنة بين موضع توزيعين أو أكثر إلا في الحالات التي يكون فيها الشكل العام لهذه التوزيعات متشابه تقريباً . وعلى ذلك لا بد من التعرف على هذه الاشكال كمتطلب سابق لاستخدام المتوسطات في المقارنة بينها .

حـ ـ لا يعتبر المتوسط وحده كافياً لوصف توزيع تكراري لأنه لا يقدم أية معلومات عن كيفية اختلاف المشاهدات التي يحسب منها المتوسط فيما بينها .

وتتطلب التطبيقات الاحصائية الاعتماد على انواع متعددة من المتوسطات . ويرجع ذلك إلى الأنواع المختلفة للبيانات المستخدمة وطبيعة التوزيعات الناتجة ، وما قد يتطلبه ذلك من أساليب تحليلية مناسبة . وسنناقش فيما بعد كيفية حساب وتفسير هذه المتوسطات المختلفة . وتجدر الاشارة إلى

أن الملاحظات السابقة حول قياس موضع التوزيع التكراري تصدق على جميع هذه المتوسطات .

٣ ـ مقاييس التشتت (أو الاتساع)

سبقت الإشارة إلى أنه لا يمكن الاعتماد على المتوسطات فقط لوصف التوزيع التكراري ، وإنما لا بد بالإضافة إلى ذلك من دراسة درجة اختلاف المشاهدات فيما بينها . ويمكن أن يتم ذلك بحساب مقياس مناسب لدرجة التشتت أو الاتساع في البيانات . وهناك عدد من مقاييس التشتت تستخدم في التطبيقات الإحصائية المختلفة ، التي تشمل بعض أو كل ما يلي :

أ ـ استخدام مقياس التشتت كأداة أساسية من أدوات وصف التوزيع التكراري ، حيث يدل المقياس على درجة التركز في المشاهدات وبالتالي يمكن تحديد ما إذا كان المتوسط مقياساً جيداً للنزعة المركزية في التوزيع . إذ كلما زادت درجة التشتت كلما قلت جودة المتوسط .

ب ـ استخدام مقاييس التشتت للمقارنة بين درجة تركز المشاهدات في توزيعين أو أكثر . وقد سبقت الإشارة إلى أن الاختلاف في تشتت التوزيعات المختلفة يؤدي إلى عدم إمكانية الاعتماد على المتوسطات في المقارنة بين موضع هذه التوزيعات .

حـ ـ يستخدم مقياس التشتت في دراسة طبيعة ومسببات هذا التشتت من أجل محاولة التحكم في مستواه . مثال ذلك التطبيقات الصحية التي تشمل قياس التشتت في درجات حرارة المريض أو في عدد نبضات قلبه أو في مستوى ضغط دمه كمتطلب سابق لإعطائه الدواء المناسب لضبط هذا التشتت .

ء ـ يعتمد على مقياس التشتت في تحديد نوع العينة التي تستخدم في إجراء دراسة ما وفي تحديد الحجم المطلوب لهذه العينة . ذلك أنه كلما كانت مفردات المجتمع غير متجانسة (أي درجة التشتت كبيرة) كلما تطلب

ذلك استخدام عينة كبيرة حتى تمثل فيها جميع أوجه الاختلاف في المجتمع . وقد يلجأ الباحث في مثل هذه الحالة إلى تقسيم المجتمع إلى طبقات واستخدام عينة طبقية لإجراء الدراسة .

ويعتبر المدى أبسط مقاييس التشتت. أما الانحراف المعياري فهو اكثر هذه المقاييس استخداماً في التطبيقات العملية . ويقيس الانحراف المعياري درجة الاختلاف بين المشاهدات ووسطها الحسابي ، حيث تكبر قيمة الانحراف المعياري كلما تقل درجة تركز المشاهدات حول وسطها الحسابي . وسوف نناقش بالتفصيل فيما بعد مقاييس التشتت المختلفة وكيفية حسابها واستخداماتها في وصف وتحليل البيانات الاحصائية .

٤ ـ مقاييس الالتواء

يقصد بالالتواء عدم التماثل في شكل التوزيع التكراري . ويكون الالتواء لليسار اذا كان هناك عدداً قليلاً من المشاهدات (التكرارات) ذات قيم صغيرة نسبياً مما يجعل الذيل الأيسر للمنحنى أكثر امتداداً من ذيله الأيمن ، ويكون الالتواء لليمين اذا كان هناك عدداً قليلاً من التكرارات ذات القيم الكبيرة نسبياً . انظر شكل (٣٨) صفحة (١٦٠) . وتشمل دراسة الالتواء في التوزيع التعرف على درجة الالتواء من ناحية وعلى اتجاه هذا الالتواء من ناحية أخرى .

ويعتبر وصف الالتواء في التوزيع التكراري أمراً هاماً للأسباب الآتية :

أ ـ أن درجة واتجاه الالتواء هي من السمات الأساسية للتوزيع التكراري ، وهي سمات لا بد من الإحاطة بها عند محاولة وصف نمط الاختلاف في البيانات .

ب ـ يؤثر درجة واتجاه الالتواء في كفاءة المقاييس المستخدمة لقياس الموضع والتشتت وفي كيفية تفسير هذه المقاييس . فمثلاً ، إذا كان هناك خمسة أسر دخولها الشهرية هي ٥٠٠٠ ، ٧٠٠٠ ، ٦٠٠٠ ، ٨٠٠٠ ، ٤٠٠٠٠

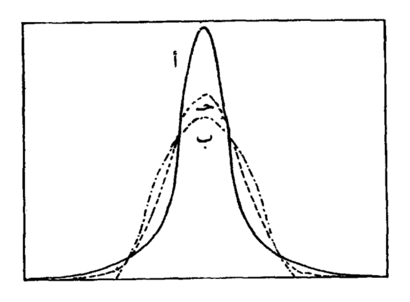
درهم على الترتيب ، حيث يلاحظ وجود التواء حاد لليمين في هذه البيانات ، يكون الوسط الحسابي لدخل الأسرة هو $\frac{٦٦٠٠٠}{٥}$ أي ١٣٢٠٠ درهم ، وهي قيمة لا تعبر عن النزعة المركزية في البيانات نتيجة تأثرها الواضح بالتواء التوزيع في اتجاه القيمة المتطرفة ٤٠٠٠٠ .

ويمكن قياس الالتواء من خلال حساب معاملات مناسبة ، سنشير إلى بعضها فيما بعد . كذلك يمكن الاكتفاء بوصف غير كمي للشكل العام للالتواء كما يبدو للباحث من دراسة جدول التوزيع أو الرسم البياني المناظر له .

٥ ـ مقاييس التفرطح

تعتبر درجة التفرطح إحدى السمات الهامة للشكل العام للتوزيع التكراري ، وقد أشير إلى ذلك بإيجاز شديد في شكل (٣٩) صفحة (١٦١) . ويقصد بدرجة التفرطح نمط التوزيع التكراري كما يبدو من درجة تركز المشاهدات حول المركز ومن شكل ذيلي التوزيع . ويؤخذ المنحنى الطبيعي كمعيار لقياس التفرطح في التوزيعات التكرارية . وفي هذا الصدد ، يعتبر هذا المنحنى معتدل التفرطح .

يوضح شكل (٣) أمثلة لأنواع التفرطح في المنحنيات التكرارية ، حيث يلاحظ أن المنحنى جـ معتدل التفرطح . إذا كانت المشاهدات مركزة حول مركز المنحنى وكان معدل التناقص من قمة المنحنى إلى ذيله أقل من المعدل المناظر للمنحنى الطبيعي بحيث يكون هناك تركيز واضح للمشاهدات في ذيلي هذا المنحنى كما يبدو بالنسبة للمنحنى (أ) في شكل (٣) ، فإن المنحنى الناتج يكون منحنى مدبباً ، وهو منحنى أقل تفرطحاً من المنحنى الطبيعي . أما إذا كانت المشاهدات تملأ جسم المنحنى بشكل منتظم وكان معدل التناقص نحو ذيليه أعلى من نظيره للمنحنى المعتاد بحيث يكون الذيلين قصيرين على النحو الذي يظهر في المنحنى (ب) في شكل (٣) ، فإن المنحنى الناتج يكون منحنى مفرطحاً .



شكل (٣) : أمثلة لمنحنيات تكرارية :
(أ) منحنى مدبب .
(ب) منحنى مفرطح .
(حـ) منحنى معتدل .

ويمكن حساب مقاييس عددية مناسبة لوصف درجة التفرطح في التوزيع التكراري . ولما كان انشاء هذه المقاييس يتطلب معرفة رياضية متقدمة ، فإنه سيكتفى في هذا الكتاب بالإشارة إلى الاعتماد على وصف درجة التفرطح في التوزيع وصفاً غير كمي بناءاً على دراسة الجدول أو الرسم البياني المناظر للتوزيع .

٦ ـ المقاييس الوصفية لبيانات المجتمع وبيانات العينة

سبقت الإشارة إلى أن هدف العمل الإحصائي في معظم الأحيان يتمثل في استخدام أساليب الاستنتاج الإحصائي للتعرف على خصائص المجتمع بناءاً على نتائج عينة من هذا المجتمع . وتسمى كل خاصية من خصائص المجتمع معلمة بينما تسمى كل خاصية من خصائص العينة إحصاءاً . وتعتمد أساليب الاستنتاج الإحصائي على استخدام احصاءات العينة لدراسة معالم المجتمع ، وقد يأخذ ذلك شكل محاولة تقدير هذه المعالم أو شكل اختبار

فروض تتعلق بهذه المعالم . وقد يتطلب الأمر وجود اختلافات بسيطة في كيفية حساب كل من مقاييس العينة ومقاييس المجتمع وذلك لأسباب ستناقش في حينها .

جرت العادة في ضوء ذلك ، على التفرقة بين المقاييس الوصفية لبيانات مجتمع والمقاييس الوصفية لبيانات عينة ، واستخدام رموز مختلفة لتمثيل كل منهما منعاً لحدوث لبس . ويتفق الاحصائيون في هذا الصدد على استخدام الرموز اليونانية (α ، β، σ ، μ، . . .) لتمثيل معالم المجتمع ، على حين تستخدم الرموز العادية (أ ، ب، ع، س، . . .) لتمثيل احصاءات العينة . وسوف نناقش في الأبواب التالية كيفية حساب المعالم والإحصاءات المختلفة ، مع توضيح أسباب الاختلافات في طرق الحساب بينها إن وجدت .

وتجدر الإشارة إلى أن الاختلافات في اسلوب تحليل بيانات المجتمع وأسلوب تحليل بيانات العينة لم تناقش عند انشاء الجداول التكرارية أو رسم الأشكال البيانية وذلك لأن أساليب العمل لا تختلف في هذه الحالة . هذا بالإضافة إلى ما ذكر سابقاً من أن الجداول والرسوم البيانية لا تصلح لأغراض الاستنتاج الإحصائي وإنما لا بد من تلخيصها أولاً بالاعتماد على عدد من المقاييس الوصفية التي يمكن استخدامها لأغراض العمل التحليلي .

يتعرض الباب السادس لمقاييس الموضع ، بينما تناقش مقاييس التشتت في الباب السابع . وسوف يشار عرضاً إلى مقاييس الالتواء والتفرطح في الأماكن الملائمة في هذين البابين . ويعتمد حساب هذه المقاييس على استخدام المشاهدات في عمليات حسابية ورياضية متعددة . وتتضمن هذه العمليات استخدامات واسعة لأساليب الجمع الرياضي . ونناقش هذه الأساليب في الجزء الباقي من هذا الباب .

٧ ـ علامة المجموع

يسهل اجراء العمليات الحسابية باستخدام الرموز للتعبير عن البيانات الإحصائية . وفي هذا الصدد يمكن أن يرمز للمتغير محل الدراسة برمز ما وليكن س (يمكن أيضاً استخدام رمز آخر مثل ص أو ع أو . . . الخ) . إذا كان هناك مشاهدات عن هذا المتغير عددها ن مثلاً فإن قيمة المشاهدة الأولى يمكن أن تكتب $س_{١}$ ويقصد بذلك قيمة المتغير س للمفردة رقم ١ ، كما تكتب المشاهدة الثانية على الشكل $س_{٢}$ والثالثة $س_{٣}$ ، . . . وهكذا حتى المشاهدة الأخيرة فتكتب $س_{ن}$. فمثلاً إذا كانت درجات خمسة من الطلبة في امتحان الإحصاء الأخير هي ٧ ، ٩ ، ١٠ ، ٨ ، ٤ فإن س هي درجة الطالب، ن = ٥ ، $س_{١} = ٧$، $س_{٢} = ٩$، $س_{٣} = ١٠$، $س_{٤} = ٨$، $س_{٥} = ٤$.

إذا أردنا الحصول على مجموع هذه الدرجات ، فيمكن كتابة ذلك على الشكل :

$$س_{١} + س_{٢} + س_{٣} + س_{٤} + س_{٥} .$$

وبشكل عام يكون مجموع ن من المشاهدات على الشكل :

$$س_{١} + س_{٢} + . . . + س_{ن}$$

حيث تدل النقط « . . . » على المشاهدات التي لم تكتب اختصاراً للمجهود . وقد جرت العادة على كتابة هذا المجموع الأخير في شكل مختصر على الشكل :

$$\sum_{ر=١}^{ن} س_{ر}$$

ويقرأ ذلك « مجموع قيم $س_{ر}$ ابتداءاً من ر = ١ حتى ر = ن » . فمثلاً $\sum_{ر=١}^{٥} س_{ر}$ يعني جمع قيم س ابتداءاً من $س_{١}$ حتى $س_{٥}$. ويلاحظ أن ر ترمز لرقم المشاهدة وأن « ر=١» تحت علامة المجموع تعني بدأ الجمع ابتداء من المشاهدة رقم ١ ، كذلك فإن ن فوق علامة المجموع يعني الاستمرار في الجمع حتى المشاهدة رقم ن. وعلى ذلك يمكن استخدام علامة المجموع

للتعبير عن المجاميع الجزئية فمثلاً :

$$\underset{ر=٣}{\overset{٥}{مجـ}} س_ر = س_٣ + س_٤ + س_٥$$ وهكذا .

وتجدر الإشارة إلى أنه إذا كان واضحاً من سياق الحديث أن المجموع يتم لجميع المشاهدات فيمكن كتابة ذلك ببساطة على الشكل مجـ س ، على أن يكون مفهوماً أن ذلك يعني مجموع جميع المشاهدات عن المتغير س .

ويمكن استخدام علامة المجموع بشكل عام ، فمثلاً إذا كانت هناك مشاهدات عن متغيرين س ، ص لمجموعة من المفردات عددها ن فإن المشاهدات في هذه الحالة تكون $س_١$ ، $س_٢$ ، $س_٣$ ، . . . ، $س_ن$ ؛ $ص_١$ ، $ص_٢$ ، $ص_٣$ ، . . . ، $ص_ن$. ويمكن الحديث عن مجاميع على الشكل : $\underset{ر=١}{\overset{ن}{مجـ}} س_ر . ص_ر$ ويقصد بذلك مجموع حاصل ضرب قيم س ، ص ابتداء من قيم المفردة الأولى حتى قيم المفردة رقم ن أي أن :

$$\underset{ر=١}{\overset{ن}{مجـ}} س_ر ص_ر = س_١ ص_١ + س_٢ ص_٢ + س_٣ ص_٣ + . . . + س_ن ص_ن .$$

وكقاعدة عامة عند استخدام علامة المجموع ، يتم حساب قيمة التعبير الجبري الذي يتبع العلامة لكل قيمة من القيم المطلوبة للدليل ر ثم يجمع الناتج . فمثلاً إذا أريد حساب المجموع $\underset{ر=٣}{\overset{٤}{مجـ}} (س_ر^٢ - ص_ر)^٣$ ، فإن ذلك يعني حساب قيمة $(س_٣^٢ - ص_٣)^٣$ وقيمة $(س_٤^٢ - ص_٤)$ ثم جمع المقدارين . كذلك فإن المقدار مجـ س ص يختلف عن المقدار (مجـ س) (مجـ ص) لأن المقدار الأول يمثل مجموع حواصل الضرب على حين يمثل المقدار الثاني حاصل ضرب المجاميع .

هناك عدة قواعد جبرية للتعامل مع علامة المجموع ، يؤدي استخدامها إلى تسهيل العمليات الحسابية في كثير من الحالات . هذه القواعد هي :

القاعدة الأولى : إذا كـان هناك مشاهـدات عن متغيرين س ، ص لمجموعة من المفردات عددها ن فإن حاصل جمع مجموع المشاهـدات لكل مفردة يساوي حاصل جمع مجموع المشاهدات لكل متغير أي أن :

$$\sum_{ر=1}^{ن} (س_ر + ص_ر) = \sum_{ر=1}^{ن} س_ر + \sum_{ر=1}^{ن} ص_ر$$

ويمكن اثبات ذلك بسهولة اذا لاحظنا أن :

$$\sum_{ر=1}^{ن}(س_ر + ص_ر) = (س_1 + ص_1) + (س_2 + ص_2) + \ldots + (س_ن + ص_ن)$$
$$= (س_1 + س_2 + \ldots + س_ن) + (ص_1 + ص_2 + \ldots + ص_ن)$$
$$= \sum_{ر=1}^{ن} س_ر + \sum_{ر=1}^{ن} ص_ر$$

القاعدة الثانية : إذا جمعت قيمة ثابتة أ عدداً من المرات فإن حاصل الجمع يساوي عدد مرات الجمع مضروباً في القيمة أي أن :

$$\sum_{ر=1}^{ن} أ = ن أ$$

ويتضح ذلك إذا تذكرنا أن القيمة الثابتة أ هي قيمة لا تتغير بتغيـر قيمة الدليل ر وبالتالي فإن :

$$\sum_{ر=1}^{ن} أ = أ + أ + أ + \ldots + أ = ن أ .$$

القاعدة الثالثة : إذا أضيف مقدار ثابت لكل مشاهدة فإن مجموع القيم الناتجة يساوي مجموع المشاهدات الأصلية مضافاً إليه حاصل ضرب المقدار الثابت في عدد المشاهدات أي أن :

$$\sum_{ر=1}^{ن} (س_ر + أ) = \sum_{ر=1}^{ن} س_ر + ن أ .$$

ويتضح ذلك بتطبيق القاعـدة الأولى أولاً على المجموع $\sum_{ر=1}^{ن}$ (س ر + أ) لنحصل على

$$\sum_{ر=1}^{ن} (س_ر + أ) = \sum_{ر=1}^{ن} س_ر + \sum_{ر=1}^{ن} أ .$$

نطبق القاعدة الثانية بعد ذلك على الحد الثاني في الجانب الأيسر لنحصل على النتيجة المطلوبة .

القاعدة الرابعة : إذا ضربت كل مشاهدة في مقدار ثابت أ فإن حاصل جمع القيم الناتجة يساوي المقدار الثابت مضروباً في مجموع المشاهدات الأصلية ، أي أن :

$$\sum_{ر=1}^{ن} أ\,س_ر = أ \sum_{ر=1}^{ن} س_ر .$$

ويمكن اثبات ذلك بملاحظة أن

$$\sum_{ر=1}^{ن} أ\,س_ر = أ\,س_1 + أ\,س_2 + \cdots + أ\,س_ن$$

$$= أ\,(س_1 + س_2 + \cdots + س_ن)$$

$$= أ \sum_{ر=1}^{ن} س_ر .$$

القاعدة الخامسة :

$$\sum_{ر=1}^{ن} (أ\,س_ر + ب) = أ \sum_{ر=1}^{ن} س_ر + ن\,ب$$ ، حيث أ ، ب مقادير ثابتة .

ويمكن اثبات هذه القاعدة بتطبيق القواعد الأولى والثانية والرابعة ، ويترك ذلك كتمرين للقارىء .

مثال (١) : جمعت بيانات عن عمر الزوج (س) وعمر الزوجة (ص) في عينة من ٥ أسر فكانت البيانات التالية :

الأسرة :	١	٢	٣	٤	٥
عمر الزوج (س) :	٥٠	٢٥	٤٠	٦٠	٦٥
عمر الزوجة (ص) :	٤٠	٢٠	٤٠	٥٥	٤٥

احسب المقادير الآتية :

(أ) $مجـ \left(\frac{س}{ص}\right)$ (ب) $\frac{مجـ\ س}{مجـ\ ص}$

(حـ) $مجـ \left(\frac{س - ٣٥}{٥}\right)^{٢}$ (ى) $مجـ\ ص^{٢} - \frac{(مجـ\ ص)^{٢}}{ن}$

(هـ) $مجـ\ س\ ص - \frac{(مجـ\ س)\ (مجـ\ ص)}{ن}$

الحل :

(أ) $مجـ \left(\frac{س}{ص}\right) = \frac{س_{١}}{ص_{١}} + \frac{س_{٢}}{ص_{٢}} + \frac{س_{٣}}{ص_{٣}} + \frac{س_{٤}}{ص_{٤}} + \frac{س_{٥}}{ص_{٥}}$

$= \frac{٥٠}{٤٠} + \frac{٢٥}{٢٠} + \frac{٤٠}{٤٠} + \frac{٦٠}{٥٥} + \frac{٦٥}{٤٥}$

$= ١,٢٥ + ١,٢٥ + ١ + ١,٠٩ + ١,٤٤$

$= ٦,٠٣$

(ب) $\frac{مجـ\ س}{مجـ\ ص} = \frac{٥٠ + ٢٥ + ٤٠ + ٦٠ + ٦٥}{٤٠ + ٢٠ + ٤٠ + ٥٥ + ٤٥} = \frac{٢٤٠}{٢٠٠} = ١,٢$

(حـ) $مجـ \left(\frac{(س - ٣٥)}{٥}\right)^{٢} = \left(\frac{(٥٠ - ٣٥)}{٥}\right)^{٢} + \left(\frac{(٢٥ - ٣٥)}{٥}\right)^{٢} + \left(\frac{(٤٠ - ٣٥)}{٥}\right)^{٢}$

$+ \left(\frac{(٦٠ - ٣٥)}{٥}\right)^{٢} + \left(\frac{(٦٥ - ٣٥)}{٥}\right)^{٢}$

$= ٩ + ٤ + ١ + ٢٥ + ٣٦$

$= ٧٥$

(ى) $مجـ\ ص^{٢} - \frac{(مجـ\ ص)^{٢}}{ن} = (٤٠)^{٢} + (٢٠)^{٢} + (٤٠)^{٢} + (٥٥)^{٢} + (٤٥)^{٢}$

$$\frac{{}^{٢}(٤٥ + ٥٥ + ٤٠ + ٢٠ + ٤٠)}{٥} \;-$$

$$\frac{{}^{٢}(٢٠٠)}{٥} - ٢٠٢٥ + ٣٠٢٥ + ١٦٠٠ + ٤٠٠ + ١٦٠٠ \quad =$$

$$٨٠٠٠ - ٨٦٥٠ = \frac{٤٠٠٠٠}{٥} - ٨٦٥٠ \quad =$$

$$٦٥٠ \quad =$$

(هـ) مجـ س ص − (مجـ س) (مجـ ص) / ن

$$(٤٥ \times ٦٥) + (٥٥ \times ٦٠) + (٤٠ \times ٤٠) + (٢٠ \times ٢٥) + (٤٠ \times ٥٠) \quad =$$

$$\frac{(٤٥ + ٥٥ + ٤٠ + ٢٠ + ٤٠)(٦٥ + ٦٠ + ٤٠ + ٢٥ + ٥٠)}{٥} \;-$$

$$٢٩٢٥ + ٣٣٠٠ + ١٦٠٠ + ٥٠٠ + ٢٠٠٠ \quad =$$

$$\frac{(٢٠٠)(٢٤٠)}{٥} \;-$$

$$٩٦٠٠ - ١٠٣٢٥ \quad =$$

$$٧٢٥ \quad =$$

تمرينات

١ ـ اشرح باختصار المقصود بكل من :

(أ) مقياس الموضع .

(ب) الالتواء .

(حـ) التفرطح .

(ء) معلمة المجتمع .

٢ ـ ما هي نوع المعلومات المتاحة في التوزيع التكراري والتي لا يمكن معرفتها بمجرد حساب متوسط التوزيع ؟

٣ ـ اذكر ثلاثة أمثلة مختلفة تتوقع أن يكون فيها التوزيع التكراري ملتو ، وبين اتجاه الالتواء في كل حالة .

٤ ـ اذكر مثالاً عملياً يتوقع أن يكون فيه التوزيع التكراري مدبب ومثال آخر يتوقع أن يكون فيه التوزيع التكراري مفرطح .

٥ ـ اشرح الهدف من قياس التشتت في التوزيع التكراري . حاول أن تعطي أمثلة عملية لتوضيح اجابتك .

٦ ـ فيما يلي بيانات عن الدرجة التي حصل عليها الطالب في امتحان الاحصاء (س) والدرجة التي حصل عليها في امتحان الاقتصاد (ص) وذلك لمجموعة من ١٠ طلبة .

الطالب :	١	٢	٣	٤	٥	٦	٧	٨	٩	١٠
درجة الاحصاء (س):	١٨	١٩	١٦	١٧	١٠	١٥	١٢	١٤	١٢	١٣
درجة الاقتصاد (ص):	١٤	١٧	١٦	١٩	١٢	١٥	١٠	١٣	١٢	١٤

إحسب المقادير الآتية :

(أ) $\text{مجـ س}^{2} - \frac{(\text{مجـ س})^{2}}{\text{ن}}$ (ب) $\text{مجـ س ص} - \frac{(\text{مجـ س})\,(\text{مجـ ص})}{\text{ن}}$

(حـ) $\text{مجـ} \left(\frac{\text{س}}{\text{ص}}\right) - \frac{\text{مجـ س}}{\text{مجـ ص}}$ (ى) $\text{مجـ س}^{2} - \text{مجـ ص}^{2}$

الباب السادس

مقاييس الموضع

سبقت الإشارة إلى إمكانية استخدام مقياس للموضع في التعرف على النزعة المركزية للبيانات وفي إجراء المقارنات بين التوزيعات المختلفة . ونناقش في هذا الباب عدداً من مقاييس الموضع ، مع الاهتمام بكيفية استخدام كل من هذه المقاييس لوصف مركز التوزيع التكراري .

١ ـ المنوال

عند الحديث عن نوع السيارة الذي يفضله السكان أو عن المدرس الأكثر شعبية بين الطلبة أو عن نوع السجائر الأكثر انتشاراً بين المدخنين فإن ذلك يتضمن تحديد المنوال أو الوجه الأكثر شيوعاً للمتغير محل الدراسة .

يعرف المنوال للمتغيرات النوعية والمتغيرات المتقطعة بأنه الوجه أو القيمة الأكثر تكراراً في البيانات . ويعكس المنوال نمط النزعة المركزية في البيانات اذا كانت المشاهدات تتركز عنده بشكل واضح .

مثال (١) : فيما يلي التوزيع التكراري لنوع السيارات المسجلة في مدينة ما .

نوع السيارة	أوروبية	أمريكية	يابانية	أنواع أخرى	المجموع
عدد السيارات المسجلة	٢٥٢٠	٣٢٤٠	٩٤١٠	٥٣٤	١٥٧٠٤

لما كانت السيارات اليابانية يناظرها أعلى تكرار في الجدول (٩٤١٠)

فإن منوال نوع السيارة في هذه الحالة هو « سيارة يابانية » .

مثال (٢) : فيما يلي التوزيع التكراري لعدد أطفال الأسرة في عينة من ١٠٠ أسرة .

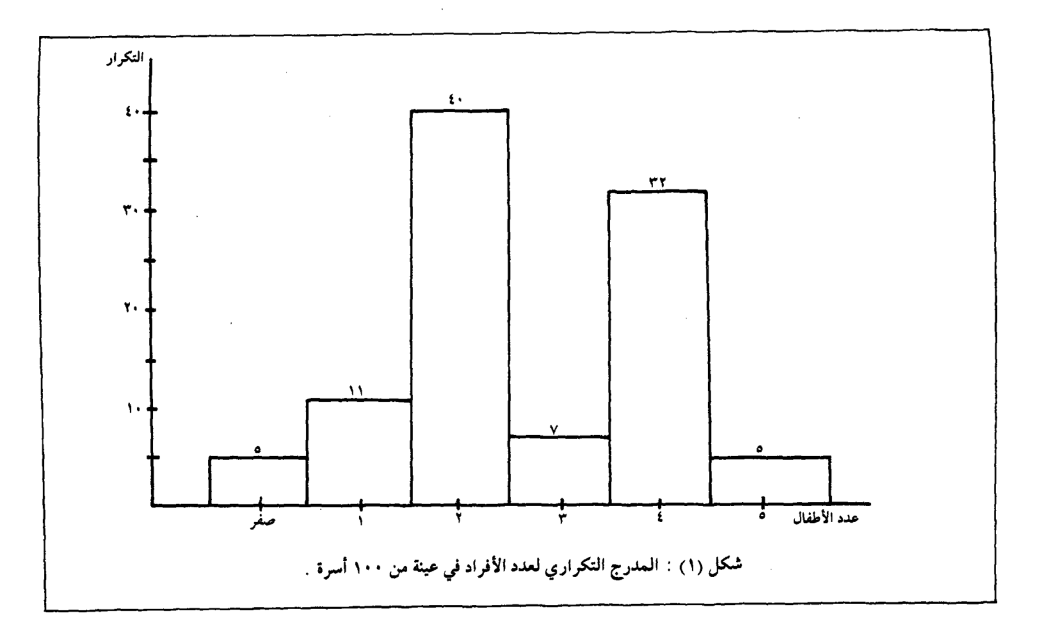
عدد الأطفال	صفر	١	٢	٣	٤	٥	المجموع
عدد الأسر	١٠	٢٣	٢٤	٤٦	١١	٥	١٠٠

ويلاحظ أن الأسرة المنوالية هنا هي الأسرة ذات الطفلين ، لأنها تناظر أعلى تكرار في الجدول وهو ٤٦ . أي أن منوال عـدد الأطفال في الأسـرة = ٢ . ويجب التأكيد على أن المنوال يكون دائماً إحدى القيم المعطاة للمتغير ، وهي قيمة تتميز بـأنها أكثر قيم هذا المتغير تكراراً . وبعبـارة أخرى ، يعتبـر المنوال مقياسـاً وصفياً للمـدرج التكراري (أو المنحنى التكـراري) المناظـر للتوزيع حيث يكون المنوال هو القيمة التي تناظر أعلى عمود في الجدول .

ولا يكون المنوال مقياساً مناسباً للموضع إلا إذا كانت هناك قيمة شائعة في البيـانات بشكـل واضح . ويتـطلب ذلك كحـد أدنى أن يكـون للمـدرج التكراري المناظر قمة واحدة متميزة . فمثـلاً ، قد يـأخذ تـوزيع عـدد أطفال الأسرة في المثال السابق الشكل التالي :

عدد الأطفال	صفر	١	٢	٣	٤	٥	المجموع
عدد الأسر	٥	١١	٤٠	٧	٣٢	٥	١٠٠

ويعـطي شكل (١) المـدرج التكراري المنـاظر ، حيث يـلاحظ وجـود قمتين للمدرج إحداها عند ٢ والأخرى عند ٤ . ويقال في هذه الحالة أن هناك منوالين للتوزيع .



شكل (١) : المدرج التكراري لعدد الأفراد في عينة من ١٠٠ أسرة.

وعادة ما يدل وجود منوالين على عدم تجانس مفردات العينة وأن البيانات تمثل خليطاً لتوزيعين يكون لكل منهما قمة واحدة . فقد يكون سبب ظهور قمتين في المثال الحالي هو وجود أسر ريفية وأسر حضرية في العينة مثلاً ، بحيث يكون توزيع عدد الأطفال في كل من هذه الأسر مشابه لما يلي ، حيث يلاحظ أن منوال عدد أطفال أسر الحضر = ٢ وأن منوال عدد أطفال أسر الريف = ٤ وأن كلاهما يفيد في وصف درجة التركز في التوزيع المناظر ولكن لا يصلح أي منهما بمفرده لوصف المجموعة الكلية للبيانات .

عدد الأطفال	صفر	١	٢	٣	٤	٥	المجموع
عدد الأسر الريفية	٥	١١	٣٥	٢	٢	١	٤٤
عدد الأسر الحضرية	صفر	صفر	٥	٥	٣٠	٤	٥٦
المجموع	٥	١١	٤٠	٧	٣٢	٥	١٠٠

ننتقل الآن إلى كيفية إيجاد المنوال للمتغيرات المتصلة مثل الطول أو الوزن أو الدخل . قد تكون المشاهدات في مثل هذه الحالات ذات قيم مختلفة نتيجة لأسلوب قياس هذه المتغيرات ، وبالتالي لا يوجد منوال بالمعنى السابق لأن التكرار المناظر لكل قيمة يساوي الواحد .

يمكن وضع البيانات في هذه الحالة في جدول تكراري باستخدام فئات مناسبة ، ويتحدد المنوال كمركز الفئة المنوالية للتوزيع . ويقصد بالفئة المنوالية تلك التي تناظر أعلى عمود في المدرج التكراري للتوزيع ، أي الفئة التي تناظر أعلى تكرار في الجدول إذا كانت فئات هذا الجدول متساوية الطول أو الفئة التي تناظر أعلى تكرار معدل إذا كانت الفئات غير متساوية الطول .

مثال (٣) : يعطي الجدول التالي التوزيع التكراري للعمر في عينة من ٤٠٠ موظف :

فئات العمر بالسنوات	٢٠ - ٢٤	٢٥ - ٢٩	٣٠ - ٣٤	٣٥ - ٣٩
عدد الموظفين	١٠	٧٠	٩٠	١١٠
فئات العمر بالسنوات	٤٠ - ٤٤	٤٥ - ٤٩	٥٠ - ٥٤	٥٥ - ٥٩
عدد الموظفين	٤٠	٤٠	٢٥	١٥
				المجموع
				٤٠٠

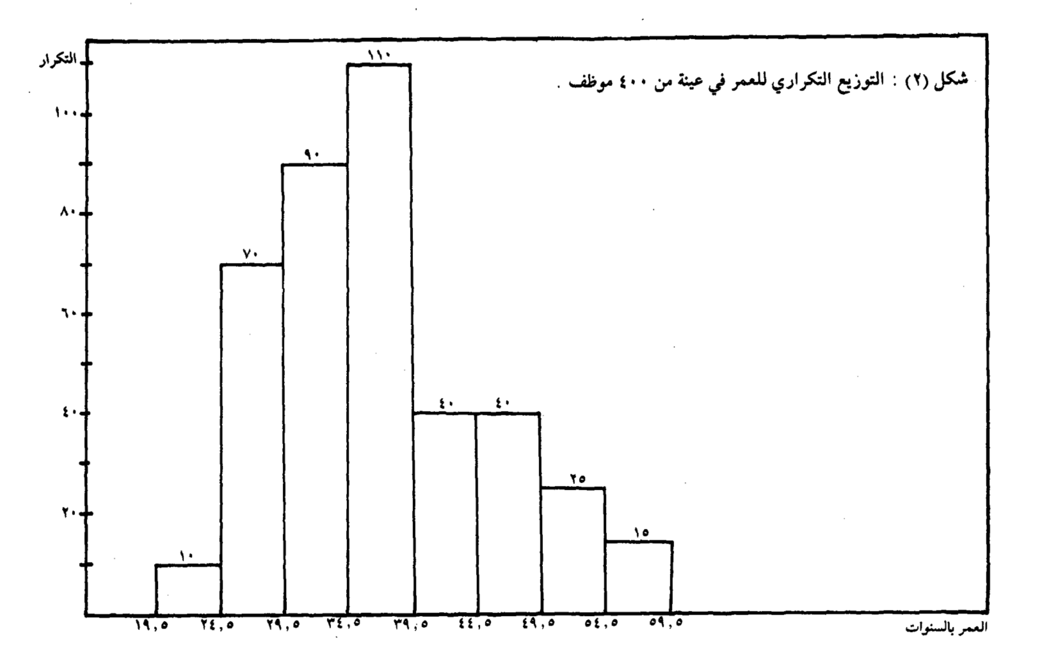
ويتضح من المدرج التكراري المناظر (شكل ٢) أن الفئة المنوالية هي الفئة « ٣٥ ـ ٣٩ » وهي الفئة التي تناظر أعلى عمود في الشكل ، ويكون المنوال هو مركز هذه الفئة أي أن :

$$\text{المنوال} = \frac{٣٥ + ٣٩}{٢} = ٣٧ \text{ عاماً .}$$

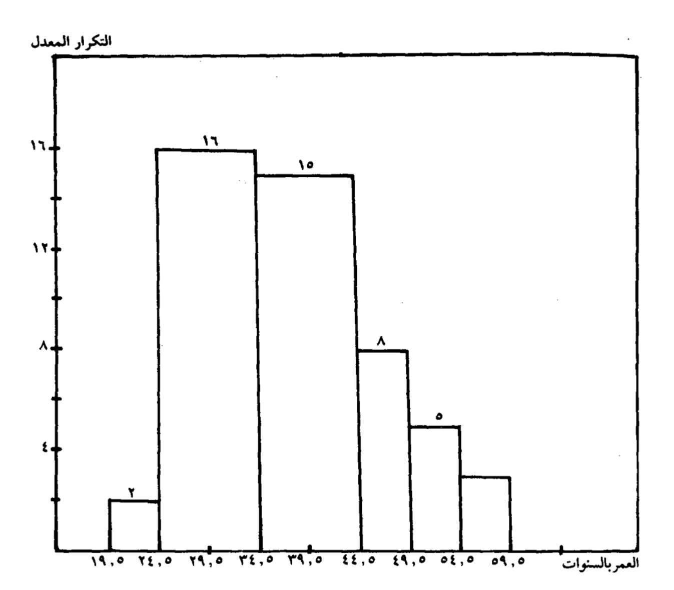
مثال (٤) : قد يعاد عرض نفس البيانات السابقة باستخدام فئات مختلفة كما يلي :

فئات العمر بالسنوات	٢٠ ـ ٢٤	٢٥ ـ ٣٤	٣٥ ـ ٤٤
عدد الموظفين	١٠	١٦٠	١٥٠
فئات العمر بالسنوات	٤٥ ـ ٤٩	٥٠ ـ ٥٤	٥٥ ـ ٥٩
عدد الموظفين	٤٠	٢٥	١٥
			المجموع
			٤٠٠

ويعطي شكل (٣) المدرج التكراري المناظر . وقد رسم هذا المدرج على أساس التكرارات المعدلة لأن فئات الجدول غير متساوية الطول . ويتضح من الشكل أن الفئة المنوالية هي الفئة « ٢٥ ـ ٣٤ » التي تناظر أعلى



شكل (٢) : التوزيع التكراري للعمر في عينة من ٤٠٠ موظف .



شكل (٣) : التوزيع التكراري للعمر في عينة من ٤٠٠ موظف .

عمود في المدرج ، ويكون المنوال هو مركز هذه الفئة أي :

$$\frac{٢٥ + ٣٤}{٢} = ٢٩ \frac{١}{٢}$$ عاماً .

يوضح هذا المثال أن القيمة الناتجة للمنوال يتأثر بكيفية اختيار الفئات التي تستخدم كأساس لانشاء التوزيع التكراري . وعلى ذلك يجب اختيار هذه الفئات بحيث يكون للمنوال معنى عند وصف التوزيع. وفي هذا الصدد ، ينصح باختيار عدد من الفئات بين ٦ ، ١٥ بحيث يكون هناك تكرارات كافية داخل كل فئة . كذلك يجب أن تكون الفئات متساوية الطول بقدر الإمكان .

وتجدر الإشارة إلى توافر عدد من الأساليب الرياضية لتحديد قيمة

المنوال داخل الفئة المنوالية ، ولكننا نفضل الاكتفاء في هذا الصدد بحساب المنوال كمركز للفئة المنوالية لعدة أسباب أهمها :

أ ـ أن المنوال في حالة المتغيرات المتصلة لا يقيس التركز عند قيمة محددة وإنما يقيس التركز داخل فئة ، وبالتالي فإن المنوال في هذه الحالة يعبر عن الفئة المنوالية ومن المنطقي الاعتماد على مركز الفئة في هذا الغرض.

ب ـ أن القيمة الناتجة للمنوال هي قيمة تقريبية على أي حال ، وقد رأينا كيف أن تغيير الفئات المستخدمة في الجدول يؤدي الى اختلاف قيمة المنوال . وبالتالي لا ينصح باستخدام أساليب رياضية معقدة قد تعطي الانطباع الخادع بارتفاع مستوى دقة النتائج . أضف الى ذلك أن المنوال مقياس وصفي بدائي لا يدخل في أية عمليات تحليلية تالية .

يلاحظ أننا استخدمنا المنوال في مثال (١) لقياس درجة التركز في بيانات نوعية . وتجدر الإشارة إلى أن المنوال ، شأنه في ذلك شأن جميع مقاييس الموضع ، يكون أكثر معنى في حالة البيانات الترتيبية والبيانات الكمية لأن هذه الأنواع من البيانات تسمح بالحديث عن النزعة المركزية . وننتقل الآن الى مناقشة مقاييس الموضع التي تستخدم لهذه الأنواع من البيانات فقط . سنبدأ بالوسيط الذي يمكن استخدامه لوصف النزعة المركزية في البيانات الترتيبية ، ثم نناقش الوسط الحسابي الذي يستخدم في وصف البيانات الكمية .

٢ ـ الوسيط

يعرف وسيط مجموعة من المشاهدات عن متغير متصل بأنه القيمة التي تقسم هذه المشاهدات إلى قسمين متساويين ، بحيث يكون عدد المشاهدات التي تقل عن أو تساوي قيمة الوسيط مساو لعدد المشاهدات التي تزيد عن أو تساوي قيمة الوسيط.

ويتم تحديد قيمة الوسيط بترتيب المشاهدات تصاعدياً ثم أخذ القيمة

التي تقع في منتصف هذه المشاهدات المرتبة .

مثال (٥) : أوجد قيمة وسيط الوزن للبيانات الآتية التي تمثل أوزان خمسة تلاميذ مقاسة بالكيلوجرام :

٥٩,٥ ، ٥٤ ، ٥٥ ، ٥٧ ، ٥٢,٥ .

الحل : (١) ترتب المشاهدات تصاعدياً :

٥٢,٥ ، ٥٤ ، ٥٥ ، ٥٧ ، ٥٩,٥

(٢) الوسيط هوالقيمة التي تقع في منتصف البيانات المرتبة ، أي أن الوسيط = ٥٥ . ويلاحظ أن هناك مشاهدتين أقل من قيمة الوسيط (هما ٥٢,٥ ، ٥٤) ومشاهدتين أكبر من قيمة الوسيط (هما ٥٧ ، ٥٩,٥) .

مثال (٦) : أوجد قيمة وسيط الوزن للبيانات الآتية التي تمثل أوزان أربعة تلاميذ مقاسة بالكيلوجرام :

٥٦,٥ ، ٥٤ ، ٥٥ ، ٥٢,٥

الحل :(١) نرتب المشاهدات تصاعدياً :

٥٢,٥، ٥٤ ، ٥٥ ، ٥٦,٥

(٢) تقع قيمة الوسيط بين ٥٤ ، ٥٥ . وقد جرت العادة على أخذ متوسط هاتين المشاهدتين :

أي : $\frac{٥٤ + ٥٥}{٢}$ = ٥٤,٥ لتمثيل قيمة الوسيط .

يتضح من هذين المثالين أنه إذا كان هناك مشاهدات عددها ن ، فإن قيمة الوسيط تكون هي القيمة ذات الرتبة $\frac{ن + ١}{٢}$ في هذه المشاهدات بعد ترتيبها . ففي المثال الأول يلاحظ أن الوسيط هو القيمة ذات الرتبة $\frac{٥ + ١}{٢}$ أي ٣ في المشاهدات المرتبة ، وفي المثال الثاني يكون الوسيط هو القيمة ذات الرتبة $\frac{٤ + ١}{٢}$ أي ٢,٥ في المشاهدات المرتبة أي القيمة التي تقع في المتوسط

بين المشاهدة ذات الرتبة ٢ والمشاهدة ذات الرتبة ٣ . ويسمى المقدار $\frac{ن + ١}{٢}$ رتبة الوسيط . ويلاحظ أن الوسيط بهـذا المعنى يمثل قيمـة وسطى للبيـانات حيث يكون عدد المشاهدات التي تزيد عن هذه القيمة مساو لعدد المشاهدات التي تقل عنها .

عند حساب الوسيط لمتغيرات مستمرة ، كما في مثالي (٥) ، (٦) ، لا يتـوقع أن يكـون هناك قيمـاً مكررة في البيـانات ، نتيجـة أسلوب قياس هـذه المتغيرات وبالتالي يمكن تفسير الوسيط في هذه الحالة بأنه القيمـة التي يقل عنها نصف المشاهـدات تقريبـاً (ويزيـد عنها نصفها الآخر) . وتـزداد جودة التقريب كلما كبر عدد المشاهدات . ولا يصدق ذلك في حالة حساب الوسيط لمتغير متقطع . فمثلاً إذا كانت البيانات الآتية تمثل عدد أفراد الأسرة في عينة من ١٠ أسر :

٢ ٢ ٢ ٣ ٣ ٣ ٣ ٤ ٤ ٤

فـإن الـوسيط في هـذه الحـالـة يسـاوي ٣ ولكن لا يـوجـد ٥٠٪ من المشاهدات أقل من هذه القيمة . ويرجع ذلك إلى إمكانية تكرار نفس القيمة في البيانات المتقطعة .

يمكن أيضاً ايجاد قيمة الوسيط عندما تكون البيانات معطاة في جـدول تكراري . ويعتمد في ذلك على التعريف التقريبي للوسيط بأنه القيمة التي يقل عنها نصف المشاهدات. ولما كان الجدول التكراري لا يفيد في اعـطاء القيم الفعلية للمشاهدات فإنه لا يمكن نظريـاً تحديـد القيمة التي يقـل عنها نصف المشاهدات بشكل دقيق تماماً . ويتم تحديد هذه القيمة عملياً بشكل تقريبي بافتراض أن المشاهدات التي تقع داخل الفئة التي تحتوي على قيمة الوسيط (الفئة الوسيطية) تكون موزعة توزيعاً منتظماً داخل هـذه الفئة . ويؤدي هـذا الافتراض عادة إلى الحصـول على تقريب جيـد لقيمة الـوسيط . وعلى ذلك يمكن تلخيص خطوات ايجاد الوسيط لتوزيع تكراري فيما يلي :

١ ـ يحدد نصف عدد المشاهدات بقسمة المجموع الكلي للمشاهدات على ٢ .

٢ ـ ترتب المشاهدات في التوزيع التكراري تصاعدياً ، ويناظر ذلك إنشاء الجدول التكراري التجميعي الصاعد (أو الهابط) .

٣ ـ تحدد الفئة الوسيطية ، وهي الفئة التي يقع داخلها قيمة الوسيط .

٤ ـ يفترض أن المشاهدات موزعة توزيعاً منتظماً داخل فئة الوسيط ، ثم يستخدم ذلك لتحديد قيمة الوسيط .

ويوضح المثال التالي خطوات العمل في هذه الحالة :

مثال (٧) : فيما يلي توزيع الدرجات التي حصل عليها ٤٠٠ طالب في امتحان الاحصاء ، احسب قيمة وسيط الدرجات في هذه البيانات .

فئات الدرجات	٤٠ ـ ٤٩	٥٠ ـ ٥٩	٦٠ ـ ٦٩	٧٠ ـ ٧٩	٨٠ ـ ٨٩	٩٠ ـ ٩٩	المجموع
عدد الطلبة	٧٠	٨٠	١٠٠	١١٠	٣٠	١٠	٤٠٠

الحل :

(١) نبدأ بتحديد نصف عدد المشاهدات ويساوي $\frac{٤٠٠}{٢} = ٢٠٠$

(٢) نكون الجدول التجميعي الصاعد :

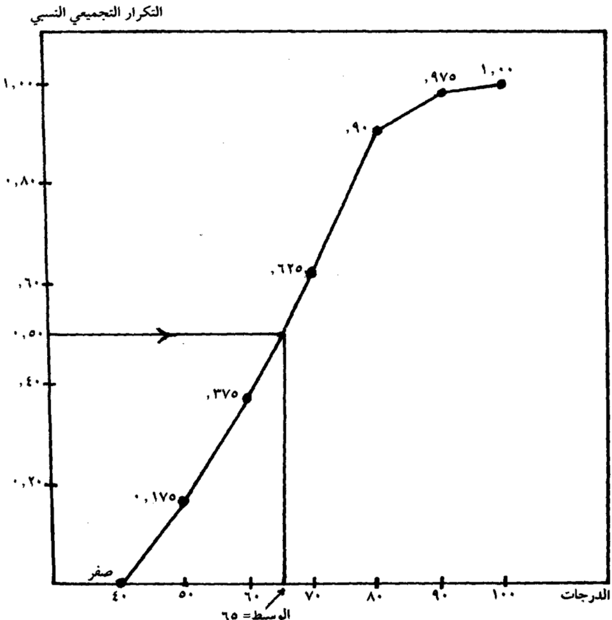
الدرجات	التكرار التجميعي الصاعد
أقل من ٤٠	صفر
أقل من ٥٠	٧٠
أقل من ٦٠	١٥٠
أقل من ٧٠	٢٥٠
أقل من ٨٠	٣٦٠
أقل من ٩٠	٣٩٠
أقل من ١٠٠	٤٠٠

(٣) نحدد الفئة الوسيطية . ولما كان الوسيط هو القيمة التي يقل عنها نصف المشاهدات (أي ٢٠٠) ، وحيث أن قيمة نصف المشاهدات

(٢٠٠) تقـع بين ١٥٠ ، ٢٥٠ في عمـود التكـرارات التجميعيـة فـإن الـوسيط يقع بين ٦٠ ، ٧٠ وبالتالي تكـون الفئة الـوسيطيـة هي « ٦٠ - ٧٠ » .

(٤) يمكن أن نكتب ما يلي بناءاً على ما سبق :

الدرجات	التكرار التجميعي
أقل من ٦٠	١٥٠
أقل من الوسيط	٢٠٠
أقل من ٧٠	٢٥٠



شكل (٤) : إيجاد الوسيط من المضلع التجميعي الصاعد .

إذا افترضنا أن المشاهدات داخل الفئة الوسيطية موزعة بانتظام فإن معنى ذلك أن الوسيط يقسم الفئة (٦٠ ـ ٧٠) بنفس النسبة التي يقسم بها المقدار ٢٠٠ الفئة (١٥٠ ـ ٢٥٠) أي أن :

$$\frac{\text{الوسيط} - ٦٠}{٧٠ - ٦٠} = \frac{٢٠٠ - ١٥٠}{٢٥٠ - ١٥٠}$$

ومنها نستنتج أن الوسيط = ٦٥ درجة .

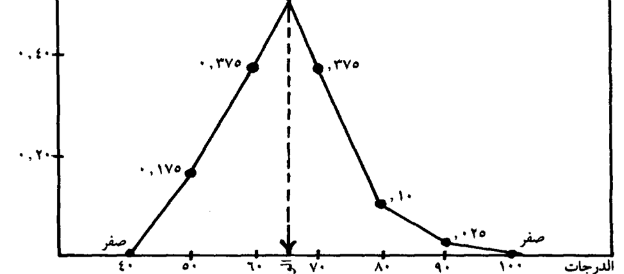
ويمكن الاعتماد على المضلع التجميعي الصاعد لايجاد قيمة الوسيط . في هذه الحالة ، يرسم المضلع التجميعي النسبي الصاعد ويكون الوسيط هو القيمة على المحور الأفقي المناظرة للتكرار النسبي ٠,٥٠ على المحور الرأسي . انظر شكل (٤) . ويلاحظ كذلك أنه إذا رسم كل من المضلع التجميعي الصاعد والمضلع التجميعي الهابط في شكل واحد ، فإن العمود الساقط من نقطة تقاطعهما يحدد قيمة الوسيط على المحور الأفقي . انظر شكل (٥) .

وتجدر الإشارة إلى أنه يمكن أيضاً استخدام المدرج التكراري لإيجاد قيمة الوسيط . ذلك أن الوسيط هو القيمة التي تقسم المدرج إلى قسمين بحيث تكون مساحة القسم الذي يقع إلى يسار الوسيط مساوياً لمساحة القسم الذي يقع إلى يمينه .

ويجب التأكيد على الصفة التقريبية لقيمة الوسيط التي تحسب من جداول التوزيعات التكرارية . ويفضل دائماً استخدام البيانات الأصلية لحساب الوسيط مباشرة كلما كان ذلك ممكناً .

٣ ـ الربيع الأول والربيع الثالث

يقسم الوسيط المشاهدات كما رأينا إلى نصفين متساويين . ويقسم الربيعين الأول والثالث والوسيط المشاهدات إلى أربع أقسام متساوية . وتفيد هذه المقاييس في وصف موضع المشاهدات البعيدة عن مركز البيانات .



شكـل (٥) : إيجاد الـوسيط من تقاطـع المضلع التجميعي الصـاعـد والمضلع التجميعي الهابط .

مثال (٨) : فيما يلي أوزان ثمـانية تـلاميذ مقـاسة لأقـرب كيلوجرام ، والمطلوب تجديد قيمة الربيع الأول وقيمة الربيع الثالث: ٥٧ ، ٥٩ ، ٥٥ ، ٥٦ ، ٥٤ ، ٥١ ، ٥٠$\frac{١}{٢}$ ، ٥٢ .

الحل : يشبه اسلوب الحل الاسلوب المتبع عند إيجاد قيمة الوسيط :

(١) نبدأ بترتيب المشاهدات ترتيباً تصاعدياً .

٥٠$\frac{١}{٢}$ ، ٥١ ، ٥٢ ، ٥٤ ، ٥٥ ، ٥٦ ، ٥٧ ، ٥٩ .

(٢) قيمـة الربيـع الأول هي ٥١$\frac{١}{٢}$ لأن هذه القيمـة تقسم المشاهـدات إلى

قسمين : الأول ويمثل ربع المشاهدات تقل قيمته عن $51\frac{1}{2}$ والثاني ويمثل ثلاث أرباع المشاهدات تزيد قيمته عن $51\frac{1}{2}$.

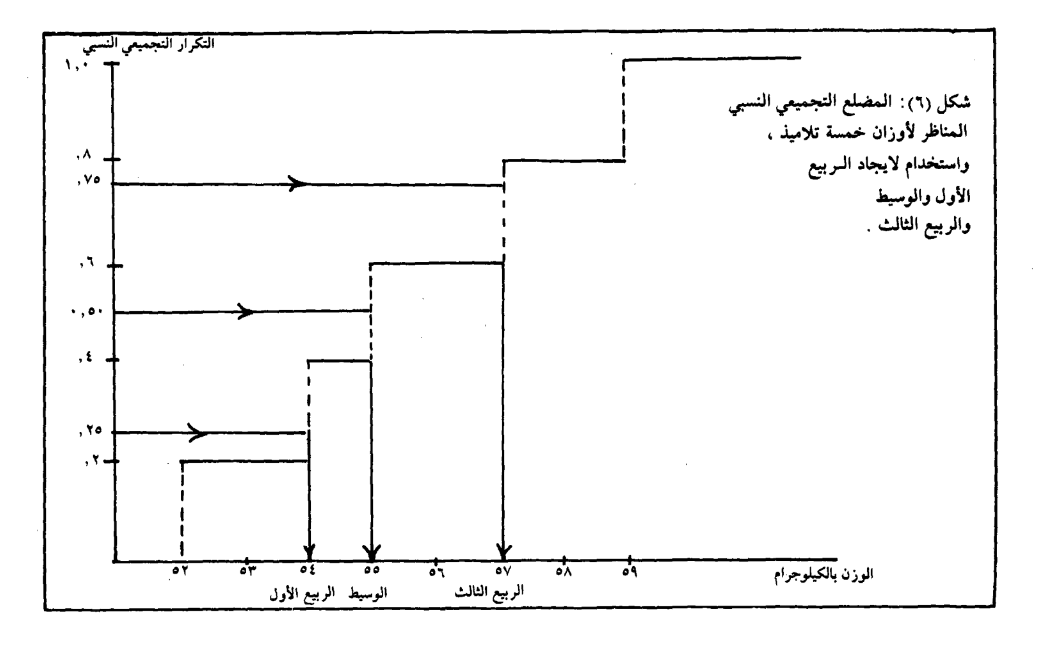
(٣) قيمة الربيع الثالث هي $56\frac{1}{2}$ لأن هناك ست مشاهدات تقل عنها ومشاهدتان تزيدان عنها .

(٤) يلاحظ أن الوسيط يعتبر الربيع الثاني ، وتكون قيمته في هذه الحالة $54\frac{1}{2}$ لأن ربعي المشاهدات تقل عن هذه القيمة بينما يزيد ربيعها الآخران عن هذه القيمة .

ويلاحظ كقاعدة عامة ، في الحالات التي يكون فيها عدد المشاهدات قليلاً ، أن إيجاد قيمة الربيع الأول وقيمة الربيع الثالث يكون أمراً سهلاً إذا كان عدد المشاهدات يقبل القسمة على ٤ . فمثلاً إذا كان هناك ثمان مشاهدات فإن قيمة الربيع الأول تقع بين المشاهدة الثانية والمشاهدة الثالثة في المشاهدات المرتبة ، وتقع قيمة الوسيط (أو الربيع الثاني) بين المشاهدة المرتبة الرابعة والمشاهدة المرتبة الخامسة ، بينما تقع قيمة الربيع الثالث بين المشاهدة المرتبة السادسة والمشاهدة المرتبة السابعة .

يمكن إيجاد قيمة الربيع الأول وقيمة الربيع الثالث في الحالة العامة التي لا يقبل فيها عدد المشاهدات القسمة على ٤ بالاعتماد على المضلع التكراري التجميعي المناظر . فمثلاً اذا كان هناك أوزان خمسة تلاميذ على الشكل : ٥٢ ، ٥٧، ٥٩، ٥٥، ٥٤، فإن التوزيع التكراري التجميعي النسبي لهذه المشاهدات يمكن أن يكتب في الصورة :

الوزن بالكيلوجرام	التكرارات التجميعية النسبية
أقل من ٥٢	صفر
أقل من ٥٤	٠,٢
أقل من ٥٥	٠,٤
أقل من ٥٧	٠,٦
اقل من ٥٩	٠,٨
اقل من حد أعلى	١,٠



شكل (٦): المضلع التجميعي النسبي المناظر لأوزان خمسة تلاميذ، واستخدام لايجاد الربيع الأول والوسيط والربيع الثالث.

ويعطي شكل (٦) المضلع التجميعي المناظر . ويلاحظ من الشكل أن القيمة على المحور الأفقي التي تناظر نسبة ٠,٢٥ من التكرارات على المحور الرأسي هي ٥٤ ، وعلى ذلك تكون قيمة الربيع الأول مساوية ٥٤ . كذلك يلاحظ أن قيمة الوسيط (الربيع الثاني) تساوي ٥٥ بينما تكون قيمة الربيع الثالث مساوية ٥٧ .

إذا اعطيت البيانات المتصلة في شكل جدول تكراري ، فإنه يمكن إيجاد قيم الربيعين باتباع نفس أسلوب إيجاد الوسيط كما يتضح من المثال التالي .

مثال (٩) : اوجد قيمة الربيع الأول وقيمة الربيع الثالث لتوزيع الدرجات التي حصل عليها ٤٠٠ طالب في مادة الإحصاء والمعطاة في المثال رقم (٧) صفحة (٢٠٩) .

الحل :

(أ) إيجاد قيمة الربيع الأول

نبدأ بتحديد قيمة ربع عدد المشاهدات ويساوي $\frac{٤٠٠}{٤}$ = ١٠٠ ، ثم نكون الجدول التجميعي الصاعد كما فعلنا في مثال (٧) صفحة (٢٠٩) .

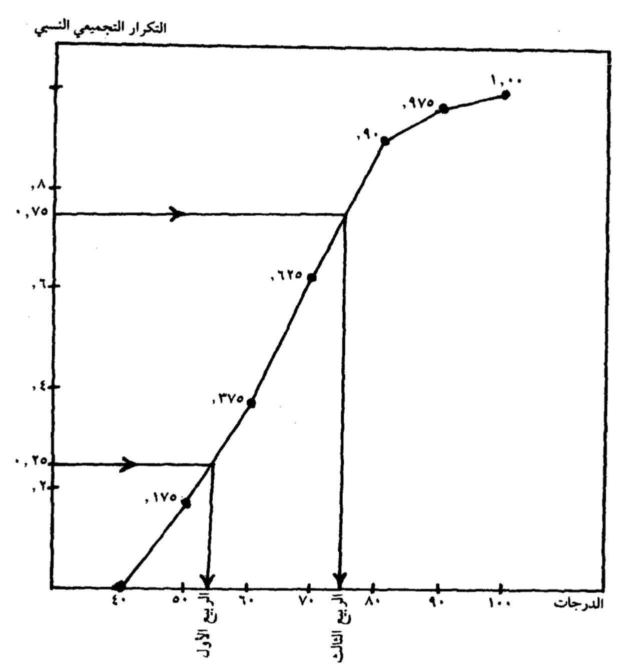
نحدد بعد ذلك فئة الربيع الأول ، لنجد أن هذه الفئة هي « ٥٠ - ٦٠ » . وبالتالي يمكن أن نكتب :

الدرجات	التكرار التجميعي
أقل من ٥٠	٧٠
أقل من الربيع الأول	١٠٠
اقل من ٦٠	١٥٠

إذا افترضنا أن المشاهدات توزع بانتظام داخل فئة الربيع الأول فإن معنى ذلك أن :

$$\frac{\text{الربيع الأول} - ٥٠}{٦٠ - ٥٠} = \frac{١٠٠ - ٧٠}{١٥٠ - ٧٠}$$

ومنها نستنتج أن قيمة الربيع الأول = ٥٣,٧٥ درجة .



شكل (٧) : إيجاد الربيع الأول والربيع الثالث من المضلع التجميعي الصاعد .

(ب) إيجاد قيمة الربيع الثالث

باتباع نفس الخطوات السابقة ، نحدد قيمة ثلاثة أرباع المشاهدات وهي ٤٠٠ × $\frac{٣}{٤}$ = ٣٠٠، لنجد أن فئة الربيع الثالث هي الفئة (٧٠ ـ ٨٠) ومن ثم :

$$\frac{\text{الربيع الثالث} - ٧٠}{٨٠ - ٧٠} = \frac{٣٠٠ - ٢٥٠}{٣٦٠ - ٢٥٠}$$

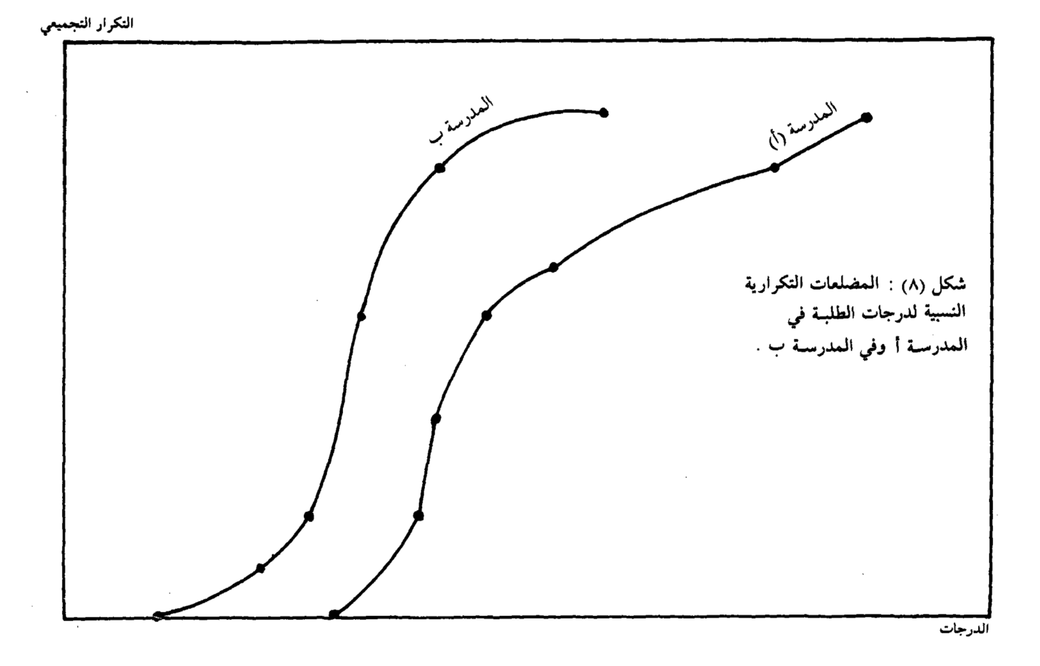
ومنها نستنتج أن قيمة الربيع الثالث = ٧٤,٥٥ درجة .

ويمكن استخدام المضلع التكراري المتجمع الصاعد لتحديد قيم الربيع الأول والربيع الثالث كما يظهر في شكل (٧) . ذلك أن الربيع الأول هو القيمة على المحور الأفقي المناظرة لتكرار تجميعي نسبي = ٠,٢٥ كما أن الربيع الثالث هو القيمة المناظرة لتكرار تجميعي نسبي يساوي ٠,٧٥ .

يعتبر الوسيط والربيعين حالات خاصة من مقاييس احصائية عامة تهدف إلى وصف نمط الترتيب في المشاهدات . فمثلاً هناك العشيرات التي تقسم المشاهدات المرتبة إلى عشرة أقسام متساوية ، ويكون العشير الثالث هو القيمة التي يقل عنها ٣٠٪ من المشاهدات ويكون العشير الخامس هو الوسيط وهكذا . كذلك هناك المييئات التي تقسم المشاهدات المرتبة إلى مائة قسم ، ويكون المييء التسعون هو القيمة التي يقل عنها ٩٠٪ من المشاهدات ويكون المييء الخمسون هو الوسيط والمييء الخامس والعشرون هو الربيع الأول وهكذا .

وتفيد هذه المقاييس في إجراء المقارنات بين المجتمعات المختلفة ، فمثلاً إذا أريد المقارنة بين مستوى الطلبة في مدرستين ، يمكن الاعتماد على العشير التاسع لدرجات الطلبة في كل منهما لإجراء هذه المقارنة . كذلك يمكن رسم المضلع التكراري التجميعي النسبي للطلبة في كل مدرسة واجراء المقارنة . فمثلاً إذا كان هذين المضلعين ممثلين في شكل (٨) فإن معنى ذلك أن جميع العشيرات والمييئات لدرجات الطلبة في المدرسة أ اكبر منها في المدرسة ب ، ويدل ذلك على أن مستوى طلبة المدرسة أ أعلى من مستوى طلبة المدرسة (ب) .

وتحسب قيم العشيرات والمييئات بنفس اسلوب حساب قيم الوسيط والربيعين وذلك بالاعتماد على التكرارات التجميعية الصاعدة . وقد يجد القارىء سهولة في الاعتماد على رسم المضلع التجميعي في هذه الحالة وقراءة قيم المقاييس من الشكل مباشرة .



شكل (٨) : المضلعات التكرارية النسبية لدرجات الطلبة في المدرسة أ وفي المدرسة ب .

٤ ـ الوسط الحسابي

يعتبر الوسط الحسابي اكثر المقاييس استخداماً لوصف البيانات الاحصائية الكمية . ويعرّف الوسط الحسابي بأنه مجموع المشاهدات مقسوماً على عددها . ويعني ذلك أن الصيغة الرياضية لحساب الوسط الحسابي للمشاهدات $س_١$ ، $س_٢$ ، . . . ، $س_ن$ والتي عددها $ن$ هي :

$$\bar{س} = \frac{مجـ س}{ن} \qquad (١)$$

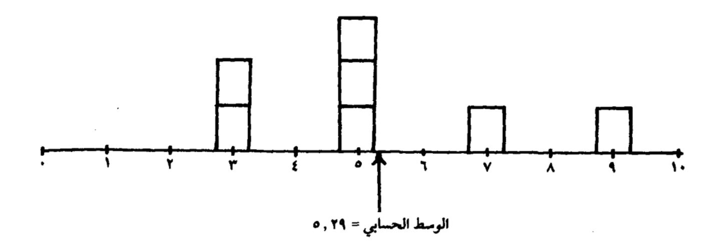
فمثلاً ، إذا كانت هناك المشاهدات التالية والتي تمثل كمية الوقود التي تستهلكها السيارة خلال يوم ما (باللتر) ، لعينة من سبع سيارات :

٥ ، ٥ ، ٣ ، ٧ ، ٩ ، ٥ ، ٣

فإن الوسط الحسابي لكمية الوقود التي تستهلكها كل من هذه السيارات يكون :

$$\bar{س} = \frac{مجـ س}{ن} = \frac{٥ + ٥ + ٣ + ٧ + ٩ + ٥ + ٣}{٧} = \frac{٣٧}{٧} = ٥,٢٩ \text{ لتراً}$$

ولفهم معنى هذه النتيجة ، تخيل أن لدينا مسطرة خشبية ، وأن المشاهدات تمثل بأثقال متساوية توضع عند قيم هذه المشاهدات على هذه المسطرة ، كما في شكل (٩) . إذا تحركت هذه المسطرة فوق سكين حاد حتى تصبح متزنة ، فإن مكان السكين يكون عند الوسط الحسابي للبيانات . ويعبر عن ذلك بأن الوسط الحسابي هو مركز الثقل للبيانات .



شكل (٩)

كمثال آخر ، فيما يلي عدد أفراد الأسرة في عينة من ١٠ أسر :

٢ ، ٤ ، ٤ ، ٣ ، ٤ ، ٣ ، ٣ ، ٣ ، ٦ ، ٣

والوسط الحسابي لعدد أفراد الأسرة فى هذه البيانات هو :

$$\bar{س} = \frac{مجـ س}{ن} = \frac{٢ + ٤ + \cdots + ٦ + ٣}{١٠} = ٣,٥ \text{فرداً}$$

ويلاحظ على الوسط الحسابي ما يلي :

أ ـ يتميز الوسط الحسابي بسهولة حسابه . ويستخدم هـذا المتوسط أسـاساً لوصف البيانات الكمية . ويكـون الوسط الحسـابي في حالـة البيانـات المتصلة إحدى القيم الممكنة للمتغيـر محل الـدراسـة ، فمثـلاً يمكن لسيارة أن تستهلك ٥,٢٩ لتراً من الوقود يومياً . أمـا في حالـة البيانـات المتقطعة ، فإن الوسط الحسابي لا يكون بالضرورة إحدى القيم الممكنة للمتغير ، مثال ذلك الحصول على القيمة ٣,٥ فرداً كمتوسط لعدد أفراد الأسرة . في هذه الحالة ، يفسر المتوسط على أساس علاقتـه بمجموع المشاهدات فإذا كان هذا المتوسط محسـوباً من عينـة من ١٠ أسر فـإن معنى ذلك أن مجموع عدد أفراد هذه الأسر يسـاوي ٣,٥ × ١٠ = ٣٥ فرداً . ويجب تبعاً لذلك اعطاء قيمة الوسط بدقة ودون تقريب .

ب ـ سبقت الإشارة إلى قواعد الدقة المتبعة في تسجيل البيانات الإحصائية . وسوف نرى فيما بعد كيف أن مستوى الدقة في الوسط الحسابي لمجموعة من المشاهدات يكون أعلى بشكل عام من مستوى دقة أي مشاهدة على حدة . لذلك ، من المتبع حساب قيمة الوسط الحسابي لمستوى دقة أعلى من مستوى دقة تسجيل المشاهدات .

حـ ـ تدخل جميع المشاهدات في العملية الحسابية لايجاد الوسط الحسابي . ويؤدي ذلك إلى ارتفاع كفاءة الوسط الحسابي بشكل عام كمقياس للنزعة المركزية في البيانات . إلا أن ذلك يؤدي من ناحية أخرى إلى تأثر الوسط الحسابي بالقيم المتطرفة التي قد توجد في المشاهدات . فمثلاً إذا كان هناك سيارة ثامنة تستهلك ٦٠ لتراً من الوقود خلال هذا اليوم في المثال الأول فيما سبق ، فإن الوسط الحسابي في هذه الحالة يصبح

$\bar{س} = \frac{٣٧ + ٦٠}{٨} = ١٢,١٣$ لتراً أي أن الوسط الحسابي قد قفز من ٥,٢٩ إلى ١٢,١٣ بسبب وجود مشاهدة شاذة عن باقي المشاهدات هي ٦٠ ، ويصبح الوسط الحسابي في هذه الحالة ممثلاً غير جيد للنزعة المركزية في البيانات .

د ـ يرمز للوسط الحسابي المحسوب من بيانات مجتمع بالرمز اليوناني μ . إذا كانت هناك مشاهدات عددها ن من جميع مفردات المجتمع فإن :

$$\mu = \frac{مجـ\ س}{ن} .$$

هـ ـ ينتج عن كون الوسط الحسابي ممثلاً لمركز الثقل في المشاهدات أن مجموع انحرافات المشاهدات عن وسطها الحسابي يساوي دائماً الصفر . أي أن :

$$مجـ\ (س - \bar{س}) = صفر$$

فمثلاً إذا كانت هناك المشاهدات ٢ ، ٦ ، ١٣ حيث وسطهم الحسابي

يساوي ٧ فإن مجـ (س - $\bar{س}$) = (٢ - ٧) + (٦ - ٧) + (١٣ - ٧) = - ٥ - ١ + ٦ = صفر ويمكن اثبات هذه العلاقة رياضياً بشكل بسيط كما يلي :

$$مجـ (س - \bar{س}) = مجـ س - مجـ \bar{س}$$

$$= مجـ س - ن \bar{س} = صفر .$$

ننتقل الآن إلى كيفية حساب الوسط الحسابي للبيانات من جدول تكراري . ويتضح كيفية ذلك من المثالين التاليين :

مثال (١٠) : يعطي الجدول الآتي التوزيع التكراري لعدد أفراد الأسرة في عينة من ١٠ أسر ، والمطلوب حساب الوسط الحسابي لعدد أفراد الأسرة في هذه البيانات .

عدد الأفراد	٢	٣	٤	٥	المجموع
عدد الأسر	١	٥	٣	١	١٠

الحل :

يلاحظ أن المتغير محل الدراسة (عدد أفراد الأسرة) متغير متقطع ، ويمكن بالتالي كتابة قيم المشاهدات التي استخدمت لإنشاء الجدول التكراري . هناك أسرة واحدة بها فردين وخمس أسر بكل منها ثلاثة أفراد وثلاث أسر بكل منها أربع أفراد وأسرة واحدة بها خمس أفراد . أي أن عدد أفراد الأسر العشرة هي :

٢ ، ٣ ، ٣ ، ٣ ، ٣ ، ٣ ، ٤ ، ٤ ، ٤ ، ٥

ويكون الوسط الحسابي لهذه المشاهدات هو

$$\bar{س} = \frac{مجـ س}{ن}$$ أي أن :

$$\bar{س} = \frac{٢ + (٣ + ٣ + ٣ + ٣ + ٣) + (٤ + ٤ + ٤) + ٥}{}$$

$$= \frac{(1 \times 2) + (5 \times 3) + (3 \times 4) + (1 \times 5)}{1 + 5 + 3 + 1} = \frac{34}{10} = 3{,}4$$ فرداً ،

حيث يلاحظ أن ايجاد الوسط الحسابي يتطلب ضرب كل قيمة من القيم المعطاة للمتغير (عدد أفراد الأسرة) في عدد مرات تكرارها (عدد الأسر) ، ثم تجمع حواصل الضرب ويقسم الناتج على مجموع التكرارات في الجدول .

ويمكن التعبير عن ذلك رياضياً كما يلي . يلاحظ أن المشاهدات العشرة $س_1$ ، $س_2$ ، . . . ، $س_{10}$ بها ٤ قيم متميزة هي ٢ ، ٣ ، ٤ ، ٥ . إذا رمزنا للقيمة الأولى بالرمز $ف_1$ وللثانية بالرمز $ف_2$ وللثالثة بالرمز $ف_3$ وللرابعة بالرمز $ف_4$ فإن $ف_1 = 2$ ، $ف_2 = 3$ ، $ف_3 = 4$ ، $ف_4 = 5$. يرمز للتكرارات المناظرة بالرمز ك أي أن $ك_1 = 1$ ، $ك_2 = 5$ ، $ك_3 = 3$ ، $ك_4 = 1$ ، ويكون الوسط الحسابي على الشكل :

$$\bar{س} = \frac{ك_1 ف_1 + ك_2 ف_2 + ك_3 ف_3 + ك_4 ف_4}{ك_1 + ك_2 + ك_3 + ك_4}$$

وبصفة عامة ، إذا كان هناك مشاهدات $س_1$ ، $س_2$ ، . . . ، $س_ن$ وكان هناك م من القيم المتميزة للمتغير محل الدراسة يرمز لها بالرموز $ف_1$ ، ف ، . . . ، $ف_م$ ، وكانت التكرارات المناظرة لهذه القيم هي $ك_1$ ، $ك_2$ ، . . . ، $ك_م$ على الترتيب فإن الصيغة الرياضية لحساب الوسط الحسابي في

$$\bar{س} = \frac{مجـ ك ف}{مجـ ك} \qquad (2)$$

مع ملاحظة أن مجـ ك تمثل العدد الكلي للمشاهدات .

مثال (١١) يعطي الجدول الآتي التوزيع التكراري للزمن الذي يستغرقه الموظف بالدقائق في الوصول الى عمله صباح يوم معين وذلك لعينة من ٢٠٠ موظف ، والمطلوب حساب قيمة الوسط الحسابي للزمن الذي يستغرقه

فئات الزمن بالدقائق	٥ ـ ٩	١٠ ـ ١٤	١٥ ـ ٢٤	٢٥ ـ ٣٤	٣٥ ـ ٤٤	المجموع
عدد الموظفين	٣٠	٧٠	٥٠	٣٠	٢٠	٢٠٠

الموظف في العينة في الوصول إلى عمله .

الحل :

تتعلق هذه البيانات بمتغير متصل هو الزمن . في هذه الحالة لا يظهر في الجدول مجموعة من قيم المتغير كما هو الحال في المتغيرات المتقطعة ، بل يعتمد الجدول على مجموعة من الفئات لقيم المتغير .

تتألف طريقة حساب الوسط الحسابي في هذه الحالة من محاولة تلخيص كل فئة من فئات الجدول بقيمة متوسطة تعبر عنها ثم استخدام هذه القيم كقيم ف في الصيغة (٢) السابق استخدامها للمتغيرات المتقطعة .

وتختار قيمة مركز كل فئة للتعبير عن القيمة المتوسطة لها . ويتضمن ذلك افتراض أن مركز كل فئة يساوي تقريباً الوسط الحسابي للمشاهدات التي تقع في هذه الفئة . وعلى ذلك فإن الوسط الحسابي المحسوب من جدول تكراري يكون تقريباً للوسط الحسابي الفعلي الذي يحسب من المشاهدات الأصلية . وتجدر الإشارة أن هذا التقريب يكون جيداً كلما كانت أطوال الفئات صغيرة وكان التوزيع التكراري أكثر تفصيلاً . ويجب التأكيد في هذا الصدد على أن انشاء التوزيع التكراري لا يعتبر متطلباً سابقاً لحساب الوسط الحسابي ، بل يفضل كقاعدة عامة استخدام المشاهدات الأصلية مباشرة كلما كان ذلك ممكناً. ويمكن اجراء الحسابات باستخدام الحاسب الآلي في المواقف التي يكون فيها عدد المشاهدات كبيراً .

يتضح من ذلك أن خطوات حساب الوسط الحسابي في هذا المثال هي :

(أ) حساب قيم مراكز الفئات ونرمز لهذه المراكز بالرمز ف .

(ب) ضرب كل قيمة من قيم مراكز الفئات في التكرار المناظر ك لنحصل على قيمة ك × ف لكل فئة .

(جـ) جمع حواصل الضرب أي ايجاد مجـ ك ف .

(د) يحسب الوسط الحسابي باستخدام الصيغة

$$\bar{س} = \frac{مجـ ك ف}{مجـ ك}$$

وتظهر هذه الخطوات في الجدول التالي :

فئات الزمن	عدد الموظفين (ك)	مركز الفئة (ف)	ك × ف
٥ - ٩	٣٠	$\frac{٥ + ٩}{٢} = ٧$	٢١٠
١٠ - ١٤	٧٠	$\frac{١٠ + ١٤}{٢} = ١٢$	٨٤٠
١٥ - ٢٤	٥٠	$\frac{١٥ + ٢٤}{٢} = ١٩,٥$	٩٧٥
٢٥ - ٣٤	٣٠	$\frac{٢٥ + ٣٤}{٢} = ٢٩,٥$	٨٨٥
٣٥ - ٤٤	٢٠	$\frac{٣٥ + ٤٤}{٢} = ٣٩,٥$	٧٩٠
المجموع	٢٠٠		٣٧٠٠

ويكون الوسط الحسابي هو : $\bar{س} = \frac{مجـ ك ف}{مجـ ك} = \frac{٣٧٠٠}{٢٠٠} = ١٨,٥$ دقيقة .

ويمكن ملاحظة الآتي :

أ - أن عدم تساوي أطوال الفئات في الجدول لا يؤثر على كيفية حساب الوسط الحسابي ، ويتضح ذلك من ملاحظة أن فئات الجدول السابق غير متساوية .

ب - يمكن استخدام التوزيع التكراري النسبي لحساب الوسط الحسابي للمشاهدات . ويتضح ذلك بكتابة الصيغة الرياضية لحساب الوسط الحسابي على الشكل :

$$\bar{س} = [مجـ \left(\frac{ك}{مجـ ك}\right) ف]$$

ولما كان $\frac{ك}{مجـ ك}$ يمثل التكرار النسبي الذي يمكن أن يرمز له بالرمز $\acute{ك}$ فإن : $\bar{س} = مجـ\ \acute{ك}\ ف$.

ويفيد الوسط الحسابي بشكل عام في اجراء المقارنات بين التوزيعات التكرارية المختلفة ، كما يتضح من المثال التالي .

مثال ١٢ : فيما يلي التوزيع التكراري لعدد الأطفال في الأسرة لكل من أسر الريف وأسر الحضر في عامي ١٩٧٠ ، ١٩٨٠ في بلد ما .

عدد الأطفال	١٩٧٠ عدد أسر الريف	١٩٧٠ عدد أسر الحضر	١٩٨٠ عدد أسر الريف	١٩٨٠ عدد أسر الحضر
صفر	٣٥٠٠	٣٤٦٥٤	٢٦٩٨	٤١٩٤٢
١	١٣٩٦	١٥٣٤٨	٧٦٢	١٧٥٢٤
٢	١١٧٤	١١٨٨٢	٦٩٤	١٦٣٧٤
٣	٨٠٤	٥١٦٢	٤٤٨	١٠١٥٠
٤	٥٠٤	١٨٣٨	٣٢٦	٤٨٨٤
٥	٢٥٢	٩٢٠	١٦٨	٢٤٩٠
٦	٢٥٢	٩٢٠	١٦٤	٢٣٩٦
المجموع	**٧٨٨٢**	**٧٠٧٢٤**	**٥٢٦٠**	**٩٥٧٦٠**

ويلاحظ أنه من الصعب اجراء المقارنات بين هذه التوزيعات الأربعة في ضوء الأرقام الكثيرة الموجودة في الجدول . قد يستخدم الوسط الحسابي في هذه الحالة لحساب متوسط كل توزيع والمقارنة بينها . وفيما يلي قيم هذه المتوسطات (ويترك حسابها كتمرين للقارىء) .

	الوسط الحسابي لعدد أطفال الأسرة	
	أسر الريف	أسر الحضر
١٩٧٠	١,٣٨٨	١,٠١٩
١٩٨٠	١,٢٥٩	١,٣٢٧

وتوضح هذه المتوسطات أن عدد أطفال أسر الريف قد قل في المتوسط بين عام ١٩٧٠ ، ١٩٨٠ على حين زاد متوسط عدد أطفال أسر الحضر خلال نفس الفترة . كذلك زاد متوسط عدد أطفال أسر الحضر عن المتوسط المناظر لأسر الريف في عام ١٩٨٠ ، وهو عكس الاتجاه الذي كان سائداً في عام ١٩٧٠ .

٥ ـ مقاييس أخرى للموضع

يعتبر الوسط الحسابي والوسيط أكثر المقاييس استخداماً لوصف النزعة المركزية في البيانات . هناك أنواع أخرى من مقاييس الموضع تستخدم في بعض المواقف الخاصة. ونناقش فيما يلي أهم هذه المقاييس .

أ ـ الوسط الحسابي المعدل

سبقت الإشارة إلى أن الوسط الحسابي يتأثر بوجود قيم شاذة في البيانات . فمثلاً ، إذا كانت هناك المشاهدات

١ ، ٢ ، ٢ ، ٤ ، ٥ ، ٧ ، ٨ ، ٩ ، ١٢ ، ١٢ ، ١٣ ، ١٢٩

فإن الوسط الحسابي يساوي ١٧ ، وهي قيمة لا تمثل النزعة المركزية في البيانات نتيجة تحيزها في اتجاه القيمة المتطرفة ١٢٩ .

ويلاحظ من ناحية أخرى أن الوسيط لا يتأثر بهذه القيمة الشاذة ، ذلك أن قيمة الوسيط في هذه الحالة تساوي ٧,٥ سواء كانت القيمة الأخيرة ١٢٩ أو أي قيمة أخرى أكبر من أو تساوي ١٣ .

للتغلب على تأثير القيم المتطرفة على الوسط الحسابي ، يمكن حساب وسط حسابي معدل (أو مؤقلم) . ويحسب هذا الوسط بترتيب المشاهدات تصاعديًا ثم حذف جزء من المشاهدات في كل جانب وأخذ الوسط الحسابي

للقيم الباقية . وتسمى نسبة البيانات التي تحذف بنسبة التعديل . وقد جرت العادة على حذف عدد متساو من المشاهدات من كل جانب .

إذا أردنا حساب وسط معدل بنسبة ٣٣٪ للبيانات السابقة فإن ذلك يعني حذف مشاهدتين من كل جانب بعد الترتيب ويكون الوسط المطلوب هو :

$$\frac{٢ + ٤ + ٥ + ٧ + ٨ + ٩ + ١٢ + ١٢}{٨} = \frac{٥٩}{٨} = ٧,٤$$

كذلك فإن الوسط المعدل بنسبة ٥٠٪ يعني حذف ثلاث مشاهدات من كل جانب ويكون الوسط المطلوب هو :

$$\frac{٤ + ٥ + ٧ + ٨ + ٩ + ١٢}{٦} = \frac{٤٥}{٦} = ٧,٥$$

ويترك تحديد نسبة التعديل للباحث في ضوء ظروف البيانات . وتجدر الإشارة إلى أنه كلما كبرت هذه النسبة كلما اقتربت قيمة المتوسط المعدل من الوسيط . ولهذا السبب يقال أن الوسط الحسابي المعدل يمثل حلاً وسطاً بين الوسط الحسابي والوسيط ويهدف الى المحافظة على مزايا كل منهما .

ب ـ منتصف المدى

يعرف منتصف المدى لمجموعة من المشاهدات بأنه الوسط الحسابي لأصغر مشاهدة وأكبر مشاهدة ، أي أن :

$$\text{منتصف المدى} = \frac{\text{اكبر قيمة + أصغر قيمة}}{٢}$$

فمثلاً ، إذا كانت هناك المشاهدات ١ ، ١ ، ٢ ، ٣ ، ٥ ، ٧ فإن

$$\text{منتصف المدى يساوي} \quad \frac{١ + ٧}{٢} = ٤ .$$

ويتميز هذا المقياس بسهولة حسابه وسهولة تفسير معناه ، ويعتبر مقياساً على درجة عالية من الكفاءة عندما يكون عدد المشاهدات قليلاً . إلا أن اعتماد منتصف المدى على أكبر قيمة واصغر قيمة يؤدي إلى تأثره الواضح بأي قيم شاذة قد توجد في البيانات .

حـ ـ الوسط الهندسي

يستخدم الوسط الهندسي أحياناً كمقياس للموضع إذا كانت البيانات تتألف من نسب . فمثلاً إذا كان سعر الوحدة من سلعة ما خلال سنوات أربع متتالية هو ٣٠,٠٠ درهم ، ٣٣,٠٠درهم ، ٣٣,٦٦درهم ، ٤١,٧٤درهم على الترتيب ، وقمنا بحساب النسبة بين سعر كل عام وسعر العام السابق عليه فإن أول نسبة هي $\frac{٣٣,٠٠}{٣٠,٠٠}$ = ١,١٠ والنسبة الثانية = ١,٠٢ والنسبة الثالثة = ١,٢٤ . يمكن الاعتماد على الوسط الهندسي كمقياس للنزعة المركزية في هذه النسب .

ويعرف الوسط الهندسي هـ لمجموعة من المشاهدات $س_١$ ، $س_٢$ ، . . . ، $س_ن$ بأنه العدد المقابل للوغاريتم للوسط الحسابي للوغاريتمات المشاهدات ، أي أن :

$$لو هـ = \frac{لو س_١ + لو س_٢ + . . . + لو س_ن}{ن}$$

فمثلاً إذا أردنا حساب الوسط الهندسي للنسب الثلاث المعطاة أعلاه ، فإن :

$$لو هـ = \frac{لو ١,١٠ + لو ١,٠٢ + لو ١,٢٤}{٣}$$

$$= \frac{,٠٤١٣٩ + ,٠٠٨٦٠ + ,٠٩٣٤٢}{٣} = ,٠٤٧٨٠$$

ويكون هـ هو العدد المقابل لهذا اللوغاريتم ، أي أن هـ = ١,١١٦ ويعاب على الوسط الهندسي صعوبة حسابه وصعوبة تفسير معناه . كذلك لا يمكن حساب هذا المتوسط إذا كانت بعض المشاهدات تأخذ قيماً صفرية أو قيماً سالبة .

ء ـ الوسط التوافقي

يستخدم الوسط التوافقي أحياناً لوصف النزعة المركزية ، إذا كانت

البيانات تتألف من معدلات للتغير . فمثلاً ، في دراسة عن تدفق حركة المرور على أحد الطرق ، لوحظ أن احدى السيارات تقطع مسافة ١٠كم بسرعة ٥٠كم/ساعة وتقطع مسافة ١٠كم أخرى بسرعة ٣٠كم/ساعة ، قد يستخدم الوسط التوافقي لحساب مقياس لمتوسط السرعتين .

يعرف الوسط التوافقي ت لمجموعة المشاهدات $س_١$ ، $س_٢$ ، . . . ، $س_ن$ بأنه مقلوب الوسط الحسابي لمقاليب المشاهدات . أي أن :

$$\frac{١}{ت} = \frac{\frac{١}{س_١} + \frac{١}{س_٢} + \ldots + \frac{١}{س_ن}}{ن}$$

ففي المثال الحالي ، $\frac{١}{ت} = \frac{\frac{١}{٥٠} + \frac{١}{٣٠}}{٢} = ٠٢٦٧,$

ويكون الوسط التوافقي ، $ت = \frac{١}{٠٢٦٧,} = ٣٧,٥$كم/ساعة.

ويلاحظ أن استخدام الوسط التوافقي ملائم في هذه الحالة لأن السيارة قطعت المسافة الكلية (٢٠كلم) في زمن قدره $\frac{١}{٥} + \frac{١}{٣} = ٥٣٣,$ ساعة وبالتالي يكون متوسط السرعة هو $\frac{٢٠}{٥٣٣,} = ٣٧,٥$.

ولا يستخدم الوسط التوافقي في كثير من التطبيقات العملية ، ولا يمكن حسابه إذا كانت هناك مشاهدات تأخذ القيمة صفر أو تأخذ قيمة سالبة .

٦ ـ ملاحظات عامة على استخدامات المتوسطات

فيما يلي بعض الملاحظات الهامة التي يجب أن تؤخذ في الاعتبار عند استخدام المتوسطات لوصف النزعة المركزية في البيانات .

أ ـ العلاقة بين الوسط الحسابي للمشاهدات ومجموع المشاهدات :

يرتبط الوسط الحسابي للمشاهدات ومجموع المشاهدات بالعلاقة $\bar{س} = \frac{مجـ س}{ن}$ ، أي أن مجـ س = ن $\bar{س}$. فمثلاً إذا كان هناك مائة أسرة وكان متوسط

عدد أفراد الأسرة يساوي ٤,٥ فرداً فإن مجموع أفراد هذه الأسر يساوي ١٠٠ × ٤,٥ = ٤٥ فرداً .

إذا كـان هناك متغيـراً متقطعـاً يأخـذ القيم واحد أو صفـر وكـان هنـاك مشـاهدات عـددها ن هى س١ ، س٢ ، . . . ، سن فـإن مجـ س في هـذه الحالة يساوي عدد المشاهدات ذات القيمة واحد في البيانات ، ويكون الوسط الحسابي للمشاهدات هو $\frac{\text{مجـ س}}{\text{ن}}$ أي عدد المشاهدات ذات القيمة واحد مقسوماً على العدد الكلي للمشاهدات . ويلاحظ أن ذلك يمثل أيضـاً نسبة حـدوث القيمة واحد في المشاهدات .

تفيد هذه النتيجة في دراسة نسبة حدوث وجـه ما لمتغيـر نوعي ، مثـل نسبة الذكور في المجتمع أو نسبة الأجانب بين طلبة الجامعة أو نسبة السيارات اليابانية إلى جملة السيارات المسجلة ، . . . الـخ . ذلك أنـه إذا كان هنـاك مشاهدات نوعية س١ ، س٢ ، . . . ، سن ، فإنه يمكن أن يرمز للمشـاهدة ذات الوجه المعين بالرمز ١ وللمشاهدات الأخرى بالرمز صفر ، ويكون حساب نسبة حدوث الوجه المعين أمر مشابه تماماً لحساب الوسط الحسابي . وبالتالي فـإن النتائـج الاحصائيـة المتعلقة بخصـائص وسلوك الوسط الحسـابي يمكن تحويرها للوصول إلى نتائج مشابهة لخصائص وسلوك النسبة .

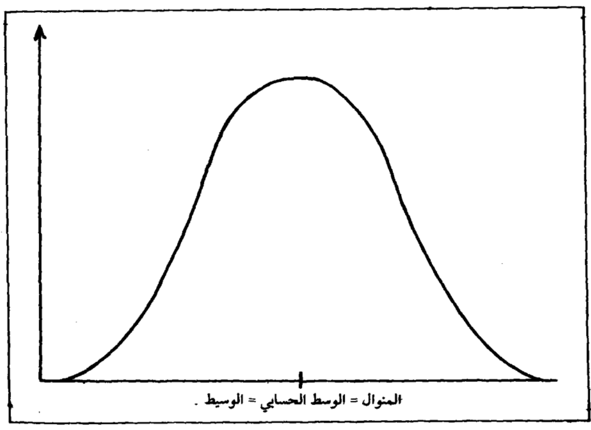
ب ـ اختيار المتوسط المناسب لوصف البيانات : يتضح فيما سبق وجود عدة أنواع من المتـوسطات تختلف فيمـا بينها من حيث الخصـائص وأسلوب الحساب . ويتطلب ذلك الاتفاق على بعض المعايير العامة للاختيار بين هذه المقاييس في التطبيقات العملية المختلفة . ويمكن القول أن تحديد المقياس المناسب للموضع في موقف ما يعتمد بصفة عامة على ثلاث عوامل أسـاسية هي :

١ ـ مفهوم القيمة المتوسطة المناسبة للهدف من الدراسة : فمثلاً عندما يقوم أحد موظفي البنوك بعد رزمة من الأوراق المـالية عـدة مرات للتـأكد من العدد الصحيح فإن المقياس المناسب يكون المنـوال . كذلـك عندمـا يكون

مجموع المشاهدات هاماً في التحليل يحسب الوسط الحسابي كما في حالـة استخدام متوسط وزن الشخص في الدراسات الخاصة بتحديد الطاقة القصوى للمصاعد .

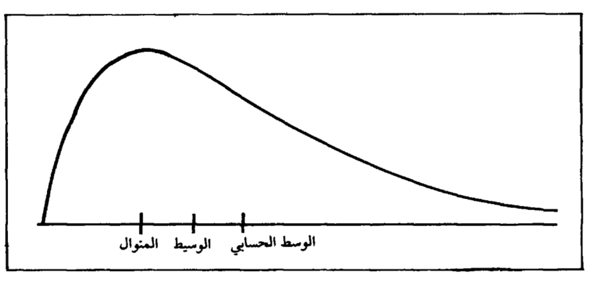
أما في الحالات التي يكون فيها الترتيب بين المشاهـدات مفيداً مثـل دراسة الدرجات التي يحصل عليها الطلبة أو الدخول التي تحصل عليها الأسر فيكون الوسيط وما شابهه هي المقاييس المناسبة .

٢ ـ **نوع البيانات المتاحـة** : يمكن أيضاً الاعتمـاد على شكل التـوزيع التكراري للبيانات لاختيار المتـوسط المناسب ، حيث يـلاحظ أن قيم الوسط الحسابي والوسيط والمنوال تكون متساوية في حالة التوزيع التكراري المتماثل ذو القمة الواحدة . ويتضح ذلك في شكل (١٠) حيث تقع نقطة توازن التوزيع (الوسط الحسابي) عند النقطة التي تناظر قمـة التوزيـع (المنوال) كمـا أن الخط العمودي الذي يصل المنحنى بهذه النقـطة يقسم المنحنى الى نصفين متماثلين في المساحة (الوسيط) .



شكل (١٠) : قيمة الوسط الحسابي وقيمة الوسيط وقيمة المنوال لتوزيع متماثل .

إذا كان التوزيع ملتوياً ، فإن المنوال يظل تحت قمة المنحنى بينما يتجه الوسط الحسابي إلى ناحية الالتواء نتيجة تأثر هذا المقياس بالقيم المتطرفة ، أما الوسيط فقد يتأثر بعدد القيم المتطرفة اذا كان هذا العدد كبيراً وبالتالي تقع قيمته بين المنوال والوسط الحسابي ، (انظر شكل (١١) ، ولاحظ أن الوسط الحسابي يمثل مركز ثقل التوزيع بينما يقسم خط الوسيط المساحة الكلية للمنحنى إلى نصفين متساويين .



شكل (١١) : قيمة الوسط الحسابي وقيمة الوسيط وقيمة المنوال لتوزيع ملتو في اتجاه اليمين .

ويعتبر الوسيط بصفة عامة مقياساً للموضع أفضل من الوسط الحسابي إذا كان التوزيع ملتوياً . ولذلك جرت العادة على حساب الوسيط لبيانات الدخل وبيانات العمر وبيانات الدرجات وبيانات فترات حضانة الأمراض ، . . . الخ لأنها جميعاً بيانات تتسم بدرجة عالية من الالتواء .

قد يكون للتوزيع التكراري قمتين . في هذه الحالة ، لا يجب الاكتفاء بمقياس واحد للموضع . إذ يجب إيجاد قيمة المنوالين، كما يمكن حساب قيمة الوسيط للدلالة على المكان الذي ينقسم عنده التوزيع إلى نصفين .

٣ ـ **الخصائص الاحصائية للمقاييس المختلفة** : ويقصد بذلك إمكانيات

استخدام هذه المقاييس لأغراض الاستنتاج الاحصائي . وقد سبقت الاشارة إلى أن المنوال لا يستخدم في هذه الأغراض . أما الوسط الحسابي فيعتبر اكثر المقاييس ملاءمة لهذا الغرض . ويرجع ذلك إلى أن الصيغة الرياضية لحسابه يسهل اخضاعها للعمليات الرياضية المختلفة ، هذا بالاضافة الى أن أخطاء المعاينة في الوسط الحسابي للعينة تكون غالباً أقل من اخطاء المعاينة في الوسيط . ويعني ذلك أن الوسط الحسابي في العينة يكون أكثر قرباً للوسط الحسابي في المجتمع .

حـ ـ المقارنة بين مجموعات البيانات غير المتجانسة : تستخدم المتوسطات للمقارنة بين النزعة المركزية في توزيعين أو اكثر . وفي هذا الصدد ، قد يؤدي الاستخدام غير الفطن لهذه المقاييس إلى الحصول على نتائج مضللة . ويحدث ذلك غالباً عندما تكون البيانات المستخدمة غير متجانسة كما يتضح من المثال الآتي .

مثال (١٣) : يعطي الجدول التالي توزيع العمال بين قطاعي الزراعة والصناعة ومتوسط أجر العامل في كل قطاع في بلدين أ ، ب :

	المدينة أ		المدينة ب	
	عدد العمال	متوسط أجر العامل	عدد العمال	متوسط أجر العامل
عمال الزراعة	٩٠,٠٠٠	٨٠ درهم	٢٠,٠٠٠	٧٠ درهم
عمال الصناعة	١٠,٠٠٠	١٠٠ درهم	٨٠,٠٠٠	٩٠ درهم
المجموع	١٠٠,٠٠٠	٨٢ درهم	١٠٠,٠٠٠	٨٦ درهم

(المصدر : بيانات فرضية)

إذا حسب متوسط اجور العمال في المدينة أ عموماً فإن القيمة الناتجة تكون :

$$\frac{١٠٠ \times ١٠٠٠٠ + ٨٠ \times ٩٠٠٠٠}{١٠٠٠٠٠} = ٨٢ \text{ درهم}$$

كذلك فإن متوسط اجور العمال في المدينة ب عموماً هو :

$$\frac{٩٠ \times ٨٠٠٠٠ + ٧٠ \times ٢٠٠٠٠}{١٠٠٠٠٠} = ٨٦ \text{ درهم}$$

إذا أجريت مقارنة بين مستوى أجور العمال في المدينتين بالاعتماد على هذه المتوسطات ، فإن هـذه المقارنـة تكون خـاطئة ومضللة . إذ يـلاحظ أن المتوسط العام للأجر في المدينة أ أقل من المتوسط العام للأجر في المدينة ب وذلك على الرغم من أن مستوى أجر عمـال الزراعـة في المدينـة أ أعلى من نظيره في المدينة ب وأن مستوى أجر عمال الصناعـة في المدينـة أ أعلى من نظيره في المدينة ب .

ويـرجع سبب هـذا التناقض إلى كيفيـة حساب المتـوسط العام لأجـور العمال في كل مدينة ، حيث يتم ضرب كل قيمة للأجـر في التكرار المنـاظر لها . وعلى ذلك يكون هذا المتوسط العام محصلة لعاملين : الأول هو مستوى الأجر والثاني هو التوزيع التكراري المناظر . ويـلاحظ أن التوزيـع التكراري للعمال في المدينة أ يختلف تماماً عن توزيع العمال في المدينة ب .

إذا أريد دراسة الاختلاف بين المدينتين في مستوى الأجر فقط ، فيجب حساب المتوسطات على أساس تـوزيع تكـراري واحد في الحـالتين . فمثلاً يمكن أخذ التوزيع التكراري في المدينة أ كأساس ، وفي هذه الحالة يكون :

$$\text{المتوسط العام للأجر في المدينة أ } = \frac{١٠٠ \times ١٠٠٠٠ + ٨٠ \times ٩٠٠٠٠}{١٠٠٠٠٠} = ٨٢ \text{ درهم}$$

$$\text{والمتوسط العام للأجر في المدينة ب } = \frac{٩٠ \times ١٠٠٠٠ + ٧٠ \times ٩٠٠٠٠}{١٠٠٠٠٠} = ٧٢ \text{درهم}$$

(على أساس توزيع العمال في أ)

حيث يلاحظ أن المقارنة بين المتوسطين سليمة وأن مستوى الأجور في المدينة أ أعلى منه في المدينة ب .

ويمكن كذلك أخذ التوزيع التكراري للعمال في المدينة ب أو أي توزيع آخر ملائم كأساس لحساب المتوسطات المقارنة . ويسمى هذا الأسلوب بأسلوب المعايرة Standardization . ويجب على القارىء أن يكون على حذر عند اجراء المقارنات الاحصائية لأن عدم استخدام اسلوب المعايرة قد يؤدي إلى متوسطات مضللة .

ء ـ تغيير وحدات القياس

يلاحظ أن التمييز المعطى لمقاييس الموضع يكون هو نفس تمييز المشاهدات التي يحسب لها هذا المقياس . اذا قيس أطوال مجموعة من الطلبة بالمتر فإن الوسط الحسابي لهذه الأطوال يكون أيضاً مقاساً بالمتر . إذا حولت هذه الأطوال إلى سنتمترات فإن الوسط الحسابي يتحول إلى سنتمترات أيضاً . أي أن ضرب كل مشاهدة في ١٠٠ في هذه الحالة يترتب عليها أيضاً ضرب الوسط الحسابي لهذه المشاهدات في ١٠٠ . ويعتبر ذلك مثالاً لقاعدة عامة تنص على أنه إذا ضربت جميع المشاهدات في نفس المقدار الموجب فإن مقياس الموضع المناظر يجب أن يضرب أيضاً في نفس المقدار الموجب .

هناك قاعدة أخرى مشابهة تنص على أنه إذا أضيف (أو طرح) مقدار ثابت إلى جميع المشاهدات ، فإن هذا المقدار الثابت يضاف (أو يطرح) أيضاً إلى مقياس الموضع المناظر . وتجدر الإشارة إلى أن مثل هذه القواعد كانت تستخدم بكثرة في الماضي لتسهيل العمليات الحسابية عند إيجاد مقاييس الموضع . وقد قل الاعتماد على هذه القواعد نتيجة الانتشار الواسع للآلات الحاسبة وتوافر تسهيلات الحاسب الآلي .

تمرينات

١ ـ فيما يلي التوزيع التكراري لعدد السيارات التي تمتلكها الأسرة في عينة من ١٠٠ أسرة .

عدد السيارات	صفر	١	٢	٣	المجموع
عدد الأسر	٤٤	٢٤	٢٤	٨	١٠٠

والمطلوب :

(أ) حساب منوال عدد السيارات للأسرة .

(ب) حساب وسيط عدد السيارات للأسرة .

(حـ) حساب الوسط الحسابي لعدد السيارات للأسرة .

(ء) حساب الربيع الأول لعدد السيارات للأسرة .

(هـ) حساب قيمة منتصف المدى لعدد السيارات للأسرة .

٢ ـ في كل حالة من الحالات الآتية ، حدد ما إذا كان الوسط الحسابي أو الوسيط أو المنوال يكون أكثر ملاءمة لوصف النزعة المركزية في البيانات . احسب قيمة المقياس الذي تختاره في كل حالة ووضح سبب اختيارك لهذا المقياس .

(أ) عدد الأطفال في ١٢ أسرة :

٣ ، ٢ ، ٢ ، ٢ ، ٣ ، ٢ ، ١ ، ٢ ، ٢ ، ٢ ، ٢ ، ٢

(ب) الدخل الشهري لعشر موظفين (بالدراهم)

٨٤٠٠ ، ٨٣٠٠ ، ٨٦٠٠ ، ٧٤٠٠ ، ٧٣٠٠ ، ٩٧٠٠ ، ٨١٠٠ ، ١٧١٠٠ ، ٩١٠٠ ، ٩٣٠٠

(حـ) أطوال عشرة مواليد بالسنتمتر

٣٩ ، ٤٠ ، ٣٩ ، ٣٧ ، ٣٨ ، ٣٩ ، ٣٨ ، ٤٠ ، ٤١ ، ٤٠ .

(ء) أعمار عشرة مصابيح كهربائية (بالساعة)

١٥٠ ، ١١٠ ، ٤٤١ ، ٢١٠٠ ، ١٥٠٣ ، ١٣٠٥ ، ٢٥٧ ،
٢٧٩ ، ٢١٥ ، ٢٥٣٦ .

(هـ) عدد السيارات التي تمتلكها الأسرة في عينة من ١٠ أسر
١ ، ١ ، ٢ ، ١ ، ١ ، ١ ، ١ ، ١ ، ٢ ، ١ .

٣ ـ (أ) وردت العبارة التالية في إحدى الصحف :

«يتضح من البيانات أن الشخص المتوسط في المدينة هو مواطن متزوج عمره ٢٨ عاماً ويسكن في منزل قيمة ايجاره الشهري ثلاثة آلاف درهم مع أسرته البالغ عدد أفرادها خمسة أفراد» . ما هي أنواع المتوسطات المحسوبة في هذه العبارة ؟

(ب) كانت قيمة منوال عدد أطفال الأسرة في بلد ما يساوي الصفر . علق أحد الباحثين على ذلك بقوله أن مستوى الانجاب في هذا البلد قد انخفض انخفاضاً ملحوظاً بحيث أن غالبية الأسر لا تنجب على الاطلاق . هل تتفق مع رأي الباحث أم لا ؟ ولماذا ؟

(حـ) قام أحد الباحثين بحساب متوسط الدرجات التي حصل عليها الطلبة في امتحان ما ولاحظ أن معظم الدرجات تقل قيمتها عن هذا المتوسط . ما هو نوع المتوسط الذي حسب في هذه الحالة ؟ هل يعتبر هذا المتوسط ملائماً لوصف البيانات ؟ اشرح سبب إجابتك .

(ء) علق على مدى صحة العبارة التالية :

« عندما تحتوي مجموعة البيانات على واحدة أو أكثر من القيم الشاذة فإنه يجب استخدام الوسيط أو الوسط الحسابي المعدل لوصف النزعة المركزية في البيانات بدلاً من استخدام الوسط الحسابي » .

٤ ـ يقوم أحد الكيميائيين بدراسة الزمن اللازم لإتمام تفاعل كيميائي معين . قرر الباحث اجراء التجربة ١٢ مرة وتسجيل هذا الزمن في كل مرة

فحصل على البيانات التالية بالدقائق .

١٧	١٢	١١	١٠	١٠	٩
١٤	١١	١٠	١٠	٩	٩

(أ) احسب الوسط الحسابي لهذه البيانات وعلق على مدى ملاءمة هذا المقياس في هذه الحالة .

(ب) اذا كانت البيانات السابقة مسجلة حسب ترتيب التجربة ، هل هناك ما يدعو الى القلق من وجود أخطاء في تصميم وتنفيذ هذه التجارب ؟ وضح سبب الاجابة .

٥ ـ أخذت عينة عشوائية من ٢٠ زجاجة مياه غازية وتم قياس حجم السائل في كل زجاجة فكانت البيانات التالية (بالسم ^{٣}) .

٤٢٣	٤٢٠	٤٢٧	٤٢٦	٤٢٦
٤٢٦	٤١٨	٤٢٨	٤٢٢	٤٢٠
٤٢١	٤٢٣	٤٢٢	٤٣٠	٤١٨
٤٢٨	٤٢٦	٤٢٢	٤٢٥	٤٢١

أ ـ احسب الوسط الحسابي لحجم السائل في الزجاجة .
ب ـ احسب وسيط حجم السائل في الزجاجة .
د ـ احسب قيمة منوال حجم السائل في الزجاجة .
هـ ـ احسب قيمة منتصف المدى لحجم السائل في الزجاجة .
و ـ احسب الوسط الحسابي المعدل بنسبة ٢٠٪ لحجم السائل في الزجاجة .

ز ـ ما الذي يمكن قوله عن مدى تماثل التوزيع في هذه البيانات وذلك بناء على نتائج الحسابات السابقة ؟

٦ ـ يعمل أحد مصانع انتاج قطع غيار السيارات على ثلاث ورديات يومية . أخذت عينة عشوائية من ٩٠ قطعة من انتاج كل دورية في أحد الأيام وصنفت تبعاً لمستوى الجودة على الأوجه أ ، ب ، جـ ، د ، هـ ، و حيث تشير أ الى جودة تامة ، ب الى جودة أقل وهكذا حتى والتي تشير إلى جودة منعدمة . وفيما يلي البيانات التي جمعت :

مستوى الجودة / الدورية	أ	ب	جـ	د	هـ	و
الدورية الأولى	١٥	٢٥	٨	٦	١٨	١٨
الدورية الثانية	١٧	٢٩	٨	١٥	٧	١٤
الدورية الثالثة	٨	٢١	١٠	٢٧	١٥	٩

أ ـ ما هي الاحصاءات التي يمكن استخدامها في هذه الحالة لوصف النزعة المركزية في كل دورية من الدوريات .

ب ـ ما الذي يمكن استنتاجه عن العلاقة بين الدورية ونوعية الانتاج ؟ هل يكون الباحث متأكداً تماماً من صحة هذه البيانات ؟ وضح سبب اجابتك .

٧ ـ فيما يلي أعمار ١٦ شخصاً :

١٠	١٤	١٣	١٨	١٢	٢٢	١٤	١٩
٢٢	١٣	١٤	١٦	٣	٦	٧	١٩

أ ـ احسب قيمة الوسط الحسابي للعمر .

ب ـ احسب قيمة الوسيط للعمر .

حـ ـ ما هو اتجاه الالتواء في هذه البيانات .

و ـ احسب قيمة الوسط الحسابي المعدل بنسبة ٢٥٪ . هل تعتقد أن

هذا المتوسط يكون أصدق تمثيلاً للبيانات في هذه الحالة ؟ وضح سبب الاجابة .

٨ - في دراسة عن زمن انتظار المريض حتى يسمح له برؤية الطبيب في إحدى المستشفيات ، جمعت البيانات الآتية عن زمن انتظار٣٠ مريضاًخلال يوم معين :

٤٠	٣٠	٤٠	٣٠	٥٥	٦٠	٣٥	٥٥	٤٠
٣٥	٥	١٠	٣٥	٦٥	٣٥	٣٠	٣٠	٦٠
٣٥	٢٥	٦٥	٣٠	١٥	٤٥	٨٥	٢٥	٢٥
١٠	١٠	٣٠						

(أ) احسب كلاً من الوسط الحسابي والوسيط ومنتصف المدى لهذه البيانات ، وعلق على مدى ملاءمة كل منها لوصف النزعة المركزية .

(ب) احسب الوسط الحسابي المعدل بنسبة ٢٠٪ .

(حـ) ضع هذه البيانات في شكل جدول توزيع تكراري مناسب .

(د) استخدم الجدول الذي تحصل عليه في (حـ) لحساب الوسط الحسابي والوسيط لزمن انتظار المريض . هل تختلف القيم التي تحصل عليها عن تلك الناتجة في (أ) ؟ وضح سبب هذه الاختلافات إن وجدت .

(هـ) ما هو اتجاه الالتواء في هذه البيانات .

٩ - (أ) وردت العبارة الآتية في احدى الصحف :

« يحصل أكثر من ٦٥٪ من العمال في المدينة على أجر يقل عن المتوسط العام لأجور العمال في هذه المدينة » .

هل يمكن أن يكون ذلك صحيحاً ؟ وضح سبب اجابتك .

(ب) ترغب وزارة التربية في تحديد ما إذا كان هناك عدد كاف من طلبة المدارس الاعدادية للالتحاق بمدرسة للنابغين تـزمع الـوزارة على افتتاحها ، حيث يكون من الضروري تحديد الحد الأدنى لمجموع الدرجات للطلبة المقبولين في هـذه المدرسة . ما هـو مقياس الموضع المناسب للاستخدام في هذه الحالة .

١٠ ـ احسب قيمة كل من الربيع الأول والربيع الثالث والوسيط للبيانات التالية التي تمثل أوزان ١٧ شخصاً .

٥٥	٥٣	٥٥	٥٩	٥٨	٥٠	٦٣	٥٠	٥٠
٤٨	٥٦	٦٣	٥٤	٥٣	٥٦	٥٠	٥٥	

١١ ـ في دراسة عن أنماط الطلاق في البلاد الغربية لوحظ أنه على الرغم من أن وسيط مدة الزواج الذي ينتهي بطلاق يساوي تقريباً سبع سنوات ، فإن معظم حالات الـطلاق تقع في السنـوات الخمس الأولى للزواج أو بعد مرور فترة تتراوح بين ٢٠ ، ٢٥ من الزواج .

أ ـ ارسم منحنى تقريبي لتمثيل مدة الزواج الذي ينتهي بطلاق .
ب ـ هل يمكنك اعطاء سبب اجتماعي لوجود قمتين في هذا التوزيع ؟

١٢ ـ فيما يلي ايجار الغرفة في أحد الفنادق خلال أربع سنوات متتالية :

السنة :	١	٢	٣	٤
الإيجار بالدرهم :	٢٥٠	٣٠٠	٤٢٠	٥٣٠

(أ) احسب نسبة التغير في ايجار الغرفة من عام إلى العام الذي يليه .
(ب) احسب الوسط الهندسي لهذه النسب .

١٣ ـ قطعت احدى الطائرات المسافة بين مـدينتين والبالغـة ١٢٠٠ كم ذهابـاً بسرعة ٣٠٠ كم/ ساعة وإياباً بسرعة ٤٠٠ كم/ ساعة . احسب الوسط

التوافقي لمتوسط السرعة ذهاباً وإياباً .

١٤ ـ يعطي الجدول التالي توزيع ١٢٥ عاملاً من عمال الخدمات حسب أجرهم اليومي بالدراهم .

فئات الأجر اليومي	١٢٠ ـ ١٢٩	١٣٠ ـ ١٣٩	١٤٠ ـ ١٤٩	١٥٠ ـ ١٥٩
عدد العمال	٩	٢٠	٣٦	٣٠

فئات الأجر اليومي	١٦٠ ـ ١٦٩	١٧٠ ـ ١٧٩	١٨٠ ـ ١٨٩	المجموع
عدد العمال	١٥	١١	٤	١٢٥

(أ) احسب الوسط الحسابي والوسيط لأجر العامل .
(ب) ما هو اتجاه الالتواء في هذه البيانات .
(حـ) احسب قيمة العشير الرابع لأجر العامل وفسر معناه .
(ء) احسب قيمة المييىء الخامس والستين لأجر العامل .
(و) أوجد قيمة المنوال لهذا التوزيع .
(ز) أحسب قيمة منتصف المدى لهذا التوزيع .
(ك) احسب قيمة الوسط الحسابي المعدل بنسبة ٢٠٪ لهذه البيانات .

١٥ ـ يعطي الجدول التالي التوزيع التكراري لنفقات الانتقال للفرد في عينة من ٢٠٠ فرداً أثناء قضائهم الصيف خارج الدولة .

فئات النفقات بالدولار	١٦٠٠ ـ ١٧٩٩	١٨٠٠ ـ ١٩٩٩	٢٠٠٠ ـ ٢١٩٩
عدد الأفراد	١٢	٣١	٧٣

فئات النفقات بالدولار	٢٢٠٠ ـ ٢٣٩٩	٢٤٠٠ ـ ٢٥٩٩	٢٦٠٠ ـ ٢٧٩٩
عدد الأفراد	٥٧	١٦	٧

فئات النفقات بالدولار	٢٨٠٠ ـ ٢٩٩٩	المجموع
عدد الأفراد	٤	٢٠٠

أ ـ احسب قيمة الوسط الحسابي لهذه البيانات .
ب ـ ارسم المضلع التجميعي الصاعد واستخدمه لإيجاد قيمة الوسيط .
حـ ـ ارسم المدرج التكراري واستخدمه لإيجاد قيمة الوسيط .
هـ ـ ما هو اتجاه الالتواء في هذا التوزيع ؟
و ـ أوجد قيمة المنوال لهذه البيانات .
ز ـ احسب قيمة الوسط الحسابي المعدل بنسبة ٤٠٪ لهذه البيانات .

١٦ ـ في كل حالة من الحالات الآتية ، اذكر ما إذا كانت هناك أية مشكلات تعوق حساب الوسط الحسابي والوسيط . وضح سبب اجابتك بشكل مفصل .

(أ) توزيع الدرجات التي حصل عليها ٧١ طالباً .

فئات الدرجات	٤٠ ـ ٤٩	٥٠ ـ ٥٩	٦٠ ـ ٦٩	٧٠ ـ ٧٩	٨٠ ـ ٩٩
عدد الطلبة	٥	١٨	٢٧	١٥	٦

(ب) توزيع الأجر اليومي الذي يحصل عليه ٦٥ عاملاً .

فئات الدخل بالدراهم	اقل من ٩٠	٩٠ ـ ٩٩	١٠٠ ـ ١٠٩	١١٠ ـ ١١٩	١٢٠ فأكثر
عدد العمال	٣	١٤	٢٢	١٩	٧

(١٧) تقوم إحدى الوكالات بتسويق أربعة أنواع من أجهزة التليفزيون . فيما يلي بيانات عن سعر بيع الوحدة من كل نوع وعدد الوحدات المباعة خلال عام معين .

نوع الجهاز	سعر بيع الوحدة	عدد الوحدات المباعة
أ	٢٥٠	٦٨٠
ب	٤٠٠	١١١٠
حـ	٦٥٠	٩١٠
ء	١٠٠٠	٥٦٠
		٣٢٥٠

أ ـ احسب الوسط الحسابي والوسيط لسعر بيع الوحدة من أجهزة التليفزيون التي تقوم هذه الشركة بتسويقها خلال ذلك العام .

ب ـ اذا كانت هناك البيانات الأصلية التي استخدمت لإنشاء هذا الجدول ، واستخدمت هذه البيانات لحساب الوسط الحسابي والوسيط مباشرة فهل نحصل على نفس الاجابات التي حصلنا عليها في (أ) ؟ ولماذا ؟

١٨ ـ ترغب إحدى شركات انتاج السيارات في الاختيار بين نوعين من اطارات السيارات باستخدام المعلومات الآتية عن المسافة التي يستخدم فيها الاطار من كل نوع حتى يستهلك تماماً :

نوع الإطار	وسيط المسافة	الوسط الحسابي للمسافة
أ	٢٥ ألف ميل	٢٧ ألف ميل
ب	٢٧ ألف ميل	٢٥ ألف ميل

بافتراض أن كلا الاطارين لهما نفس السعر ، ما هو نوع الاطار الذي تقترح على الشركة شراءه ؟ وضح سبب اجابتك .

١٩ ـ في دراسة عن الصحة العامة لتلاميذ المدارس الابتدائية الذين يتراوح أعمارهم بين ٦ ، ١١ سنة، وجد أن متوسط طول التلاميذ الذين يبلغ عمرهم ١١ عاماً يساوي ١٤٧ سم وأن متوسط طول الذكور منهم يساوي ١٤٦ سم . كذلك فإن متوسط طول الذكور الذين يبلغون من العمر ٩ سنوات يساوي ١٣٥ سم .

أ ـ ما هو متوسط طول الذكور الذين يبلغون من العمر ٩ سنوات بالبوصات (البوصة = ٢,٥٤ سم) .

ب ـ هل تدل هذه النتائج على أن الذكور الذين يبلغون من العمر ١١ عاماً أكثر طولاً من الاناث الذين يبلغون من العمر ١١ عاماً .

حـ ـ قدر بشكل تقريبي متوسط طول الذكور الذين يبلغون من العمر عشرة أعوام .

٢٠ ـ للتأكد من صحة الادعاء أن الوسط الحسابي يكون عادة أكثر دقة من الوسيط لأغراض الاستنتاج الاحصائي ، قام أحد الطلبة باجراء تجربة تتمثل في رمي ثلاث زهرات نرد اثني عشر مرة متتالية . فيما يلي نتائج الزهرات في كل مرة .

٢ ، ٤ ، ٦	٥ ، ٢ ، ٣	٣ ، ١ ، ٤	١ ، ٦ ، ٢
٥ ، ٣ ، ٥	٦ ، ١ ، ٥	٥ ، ٥ ، ٢	٣ ، ٣ ، ٣
٤ ، ٥ ، ٣	٣ ، ٢ ، ١	٣ ، ٣ ، ٤	٤ ، ٥ ، ٣

أ ـ احسب قيمة الوسط الحسابي وقيمة الوسيط لنتائج كل من هذه الرميات الاثني عشر .

ب ـ كون توزيعاً تكرارياً لقيم الوسط الحسابي وتوزيعاً تكرارياً آخر لقيم الوسيط باستخدام الفئات ١,٥ ـ ٢,٥ ، ٢,٥ ـ ٣,٥ ، ٣,٥ ـ ٤,٥ ، ٤,٥ ـ ٥,٥ . (لاحظ أن استخدام هذه الفئات مباح لأنه لا يوجد قيم متداخلة بينها لأن الوسط الحسابي لثلاثة أرقام صحيحة لا يمكن أن يساوي ٢,٥ أو٣,٥ أو٤,٥) .

حـ ـ اشرح كيف أن هذه التوزيعات توضح أن الوسط الحسابي يتعرض لأخطاء معاينة تقل عن تلك التي يتعرض لها الوسيط .

٢١ ـ قام أحد أجهزة حماية المستهلك باختبار استهلاك الوقود لثلاثة أنواع من السيارات باستخدام عينة من خمس سيارات من كل نوع . فيما يلي بيانات عن عدد الأميال لكل جالون من الوقود :

النوع أ : ٢٧,٩ ، ٣٠,٤ ، ٣٠,٦ ، ٣١,٤ ، ٣١,٧

النوع ب : ٣١,٢ ، ٢٨,٧ ، ٣١,٣ ، ٢٨,٧ ، ٣١,٣

النوع حـ : ٢٨,٦ ، ٢٩,١ ، ٢٨,٥ ، ٣٢,١ ، ٢٩,٧

أ ـ اذا أرادت الشركة التي تنتج النوع أ الإعلان عن أن هذا النوع هو أفضل الأنواع استهلاكاً للوقود ، فما هو نوع المتوسط الذي يمكن أن تعتمد عليه الشركة لهذا الغرض .

ب ـ اذا أرادت الشركة التي تنتج النوع ب الادعاء أن هذا النوع هو أفضل الأنواع استهلاكاً للوقود ، فما هو نوع المتوسط الذي يمكن حسابه لهذا الغرض .

حـ ـ بين أن استخدام منتصف المدى لقياس النزعة المركزية في هذه البيانات يؤدي الى استنتاج أن النوع جـ هو أفضل الأنواع استهلاكاً للوقود .

ء ـ ما الذي تدل عليه هذه النتائج فيما يتعلق بالاعتماد على المتوسط فقط لوصف مجموعة من البيانات .

٢٢ ـ فيما يلي عدد السكان ومعدلات الوفاة في الفئات العمرية المختلفة في كل من البلد أ والبلد ب خلال عام معين .

فئات العمر	البلد أ		البلد ب	
	عدد السكان	معدل الوفاة	عدد السكان	معدل الوفاة
صفر ـ ٤	١٠٠٠	,٠١٠	٤٠٠٠	,٠٠٧
٥ ـ ٤٤	٦٠٠٠	,٠٠٥	٤٠٠٠	,٠٠٣
٤٥ فأكثر	٣٠٠٠	,٠١٠	١٢٠٠٠	,٠١٥

أ ـ احسب المتوسط العام لمعدل الوفاة في كل من المدينة أ والمدينة ب .

ب ـ هل تصلح هذه المتوسطات للمقارنة بين مستوى الوفاة في المدينتين ؟ وضح سبب اجابتك ؟

حـ ـ اذا كانت اجابتك بالنفي عن الجزء (ب) ، اقترح وسيلة مناسبة لاجراء المقارنة بين مستوى الوفاة في المدينتين .

٢٣ ـ تقوم إحدى الشركات بتسويق أربعة أنواع من الغسالات. فيما يلي بيانات عن سعر بيع الوحدة وعدد الوحدات المباعة من كل نوع في عامي ١٩٧٠ ، ١٩٨٠ .

	١٩٧٠		١٩٨٠	
	سعر بيع الوحدة	عدد الوحدات المباعة	سعر بيع الوحدة	عدد الوحدات المباعة
النوع أ	٢٠٠	٧٠٠	٣٠٠	٥٠٠٠
النوع ب	٤٠٠	١٠٠٠	٥٠٠	٣٠٠٠
النوع جـ	٦٠٠	٩٠٠	٧٠٠	١٠٠٠
النوع ء	١٠٠٠	٥٠٠	٨٠٠	٢٠٠٠

أ ـ احسب المتوسط العام لسعر بيع الغسالات المباعة عموماً في كل من عامي ١٩٧٠ ، ١٩٨٠ .

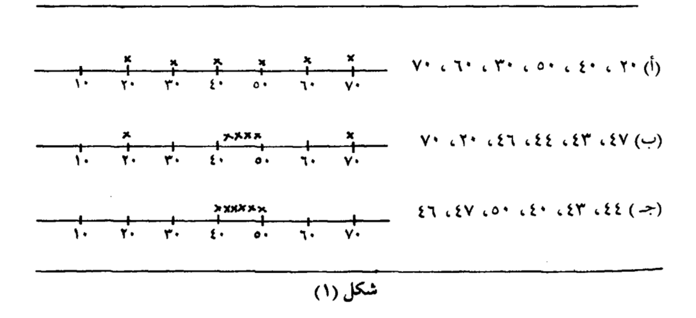
ب ـ إذا أريد المقارنة بين متوسط أسعار الغسالات في عام ١٩٧٠ ومتوسط أسعار الغسالات في عام ١٩٨٠ ، هل يمكن الاعتماد على المتوسطات المحسوبة في (أ) لأداء هذا الغرض ؟ وضح سبب اجابتك .

حـ ـ اذا كانت الاجابة عن الجزء (ب) بالنفي ، اشرح كيف يمكن انشاء متوسطات ملائمة لاجراء المقارنة المطلوبة.

الباب السابع

مقاييس التشتت

ناقشنا في الباب السابق مقاييس الموضع المختلفة التي تهتم بوصف النزعة المركزية في المشاهدات . وقد سبقت الاشارة إلى أن مقياس المركز لا يكفي كقاعدة عامة لاعطاء وصف كامل لنمط الاختلاف في البيانات . فمثلاً ، يعطي شكل (١) ثلاث مجموعات مختلفة من البيانات (تمثل كل منها أعمار ستة أشخاص) ، حيث يلاحظ أن قيم الوسط الحسابي والوسيط متساوية لجميع هذه المجموعات على الرغم من الاختلافات الواضحة بين أنماط التشتت فيها .



شكل (١)

ويعتبر نمط التشتت أحد السمات الرئيسية لتوزيع المشاهدات . وقد أشرنا سابقاً إلى أن مقياس الموضع يكون أكثر كفاءة في تمثيل البيانات عندما

تكون هذه البيانات أقل تشتتاً أو اختلافاً فيما بينها . ونناقش في هذا الباب المقاييس المختلفة للتشتت ، مع إعطاء الاهتمام لأسلوب الحساب وكيفية تفسير كل من هذه المقاييس .

١ ـ المدى

يعتبر المدى أكثر مقاييس التشتت بساطة ، ويعرف بأنه الفرق بين اكبر مشاهدة وأصغر مشاهدة في البيانات . أي أن :

المدى = اكبر مشاهدة ـ أصغر مشاهدة .

ويتميز المدى بسهولة حسابه وسهولة تفسير معناه . ويستخدم كثيراً في التطبيقات العملية ، مثال ذلك عمليات مراقبة جودة الانتاج الصناعي حيث يتحدد عادة مدى مقبول لمواصفات الانتاج ويتم فحص هذا الانتاج بشكل دوري للتأكد من عدم الخروج عن هذا المدى . كذلك يستخدم المدى في التطبيقات الطبية حيث يتفق على مدى مقبول لكيفية اداء أعضاء الجسم لوظائفها المختلفة مثل عدد نبضات القلب أو ضغط الدم أو درجة حرارة الجسم بحيث يعتبر الشخص مريضاً اذا تعدى هذا المدى . ويستخدم المدى أيضاً لوصف درجات الحرارة المتوقعة عند التنبؤ بالطقس .

ويفيد المدى في إعطاء صورة سريعة لحجم التشتت في البيانات ، وذلك لاعتماده على أصغر مشاهدة وأكبر مشاهدة فقط . وقد يؤدي هذا الاعتماد إلى إعطاء صورة مضللة للتشتت إذا كان هناك مشاهدات متطرفة أو شاذة في البيانات . فمثلاً إذا كان عدد أطفال الأسر في عينة من ثمان أسر هو صفر، ١ ، ١ ، ٢ ، ٢ ، ٣ ، ٣، ٩ فإن قيمة المدى تساوي ٩ ـ صفر = ٩، وذلك على الرغم من أن عدد أطفال معظم الأسر يتراوح بين ١ ، ٣ . كذلك يلاحظ أن قيمة المدى في مجموعتي البيانات (أ) ، (ب) في شكل (١) متساوية (٧٠ ـ ٢٠ = ٥٠) وذلك على الرغم من الاختلاف الواضح في نمط تشتتهما . لذلك ، لا ينصح باستخدام المدى كمقياس للتشتت اذا كان هناك

ما يدعو للاغتقاد بوجود قيم متطرفة في البيانات .

ويمكن حساب تقدير تقريبي للمدى إذ كانت البيانات معطاة في شكل جدول تكراري . ويحسب هذا التقدير بأخذ الفرق بين الحد الأعلى للفئة الأخيرة والحد الأدنى للفئة الأولى في الجدول .

٢ ـ المدى الربيعي

يمكن التقليل من اعتماد المدى على المشاهدات الشاذة أو المتطرفة بترتيب المشاهدات ترتيباً تصاعدياً ثم حذف جزء من المشاهدات من كل جانب وحساب قيمة المدى للمشاهدات الباقية . ويعتبر المدى الربيعي أبسط هذه المقاييس وأكثرها استخداماً في التطبيقات العملية .

ويعرف المدى الربيعي بأنه الفرق بين الربيع الثالث والربيع الأول للمشاهدات . وعلى ذلك يمكن اعتباره مدى معدلاً يصف التشتت في النصف الأوسط من المشاهدات .

مثال (١): احسب المدى الربيعي لكل من مجموعتي البيانات التالية وعلق على الناتج .

(أ) صفر، ١ ، ١ ، ٢ ، ٢ ، ٣ ، ٣ ، ٩ .

(ب) صفر، ١ ، ١ ، ٢ ، ٣ ، ٦ ، ٦ ، ٩ .

الحل :

(أ) الربيع الأول = ١ ، الربيع الثالث = ٣

المدى الربيعي = ٣ ـ ١ =٢

(ب) الربيع الأول = ١ ، الربيع الثالث = ٦

المدى الربيعي = ٦ ـ ١ = ٥

يلاحظ أن قيمة المدى في الحالتين = ٩ ـ صفر = ٩ وذلك على الرغم من وجود تشتت اكبر في مشاهدات المجموعة (ب) . ويعكس المدى الربيعي ذلك حيث تزيد قيمته للمجموعة (ب) عن قيمته للمجموعة (أ) .

مثال (٢) : يعطي الجدول الآتي التوزيع التكراري للوزن في عينة من خمسين شخصاً ، والمطلوب حساب المدى الربيعي للأوزان وتفسير معناه .

فئات الوزن بالكيلوجرام	٥٠ ـ ٥٤	٥٥ ـ ٥٩	٦٠ ـ ٦٤	٦٥ ـ ٦٩
عدد الأشخاص	٥	١٧	١٧	٧
فئات الوزن بالكيلوجرام	**٧٠ ـ ٧٤**	**٧٥ ـ ٧٩**	**٨٠ ـ ٨٤**	**المجموع**
عدد الأشخاص	٢	١	١	٥٠

الحل : نبدأ بإيجاد قيمة الربيع الأول وقيمة الربيع الثالث باستخدام إحدى الطرق التي سبق الإشارة إليها . فمثلاً يمكن حساب التكرارات التجميعية الصاعدة كما يلي :

اقل من ٥٠	صفر
اقل من ٥٥	٥
اقل من ٦٠	٢٢
اقل من ٦٥	٣٩
اقل من ٧٠	٤٦
اقل من ٧٥	٤٨
اقل من ٨٠	٤٩
اقل من ٨٥	٥٠

لإيجاد الربيع الأول ، نلاحظ أن قيمة ربع عدد المشاهدات = $\frac{٥٠}{٤}$ = ١٢,٥ ، وبالتالي يمكن أن نكتب الآتي :

اقل من ٥٥	٥
اقل من الربيع الأول	١٢,٥
اقل من ٦٠	٢٢

بحيث أن :

$$\frac{\text{الربيع الأول} - ٥٥}{٦٠ - ٥٥} = \frac{١٢,٥ - ٥}{٢٢ - ٥}$$

ومنها نجد أن الربيع الأول = ٥٧,٢١ .

كذلك ، لإيجاد الربيع الثالث ، يلاحظ أن قيمة ثلاثة أرباع عدد المشاهدات = ٥٠ × $\frac{٣}{٤}$ = ٣٧,٥ ، وبالتالي يكون :

اقل من ٦٠	٢٢
اقل من الربيع الثالث	٣٧,٥
اقل من ٦٥	٣٩

بحيث أن :

$$\frac{\text{الربيع الثالث} - ٦٠}{٦٥ - ٦٠} = \frac{٣٧,٥ - ٢٢}{٣٩ - ٢٢}$$

ويكون الربيع الثالث = ٦٤,٥٦ .

وتكون قيمة المدى الربيعي = ٦٤,٥٦ - ٥٧,٢١ = ٧,٣٥ كيلوجرام . معنى ذلك أن النصف الأوسط من اوزان هؤلاء الاشخاص ينتشر فوق مدى طوله ٧,٣٥ كيلوجرام .

ويتم عادة حساب الوسيط كمقياس للمركز عند حساب المدى الربيعي كمقياس للتشتت . ويترك للقارىء بيان أن قيمة الوسيط في المثال الحالي = ٦٠,٨٨كيلوجرام .

ويمكن أيضاً إيجاد قيمة الربيعين الأول والثالث برسم المضلع التجميعي النسبي الصاعد ، وتكون قيمة الربيع الأول هي القيمة على المحور الأفقي المناظرة لتكرار تجميعي نسبي = ٠,٢٥ كما تكون قيمة الربيع الثالث هي القيمة المناظرة لتكرار تجميعي نسبي = ٠,٧٥ ويمكن كذلك الاعتماد

على المدرج التكراري لايجاد قيمة هذه المقاييس ، اذ يكون الربيع الأول هو القيمة التي يقع ربع مساحة المدرج إلى يسارها بينما يكون الربيع الثالث هو القيمة التي يقع ربع مساحة المدرج إلى يمينها . ويترك للقارىء تطبيق هذه الطرق المختلفة للحصول على قيم الربيعين .

يمكن تعميم فكرة المدى الربيعي لإنشاء مقاييس مشابهة للتشتت . مثال ذلك المدى العشيري الذي يحسب بأخذ الفرق بين العشير التاسع والعشير الأول او المدى الخميسي ويحسب بأخذ الفرق بين العشير الثامن (الخميس الرابع) والعشير الثاني (الخميس الأول) ، وهكذا .

وتجدر الإشارة إلى إمكانية عرض قيم الربيع الأول والربيع الثالث والوسيط والمدى في شكل بياني يوضح نمط الاختلاف في البيانات . ويعتبر شكل الصندوق والشعيرات Box- and- Whiskerplot مثالاً هاماً لهذه الأشكال . وينشأ شكل الصندوق والشعيرات برسم صندوق مستطيلي يكون جانبه الأيسر مناظراً لقيمة الربيع الأول وجانبه الأيمن مناظراً لقيمة الربيع الثالث (أي أن عرض الصندوق يمثل المدى الربيعي) . يرسم خط عمودي داخل الصندوق عند قيمة الوسيط ، ثم تؤخذ شعيرة (أو خطاً أفقياً) من كل جانب من جانبي الصندوق لتغطي المدى الذي تنتشر فوقه المشاهدات . ويفيد هذا الشكل في اعطاء وصف موجز لنمط الاختلاف في البيانات يشمل معلومات عن مركزها ونمط تشتتها ودرجة التماثل أو الالتواء فيها وما إذا كان هناك مشاهدات شاذة أو متطرفة . ويتضح ذلك من الأمثلة التالية .

مثال (٣) : فيما يلي الزمن الذي استغرقه كل عداء في قطع مسافة ١٠٠ متر لمجموعة من ٢١ عداءاً (الزمن بالثواني) .

١٢,٨١ ١٤,٩٥ ١٥,٨٣ ١٥,٩٧ ١٧,٩٠ ١٨,٢٧ ١٨,٣٤ ١٩,٨٢

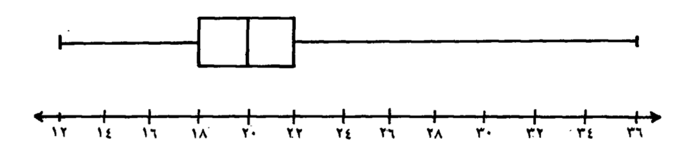
١٩,٩٤ ٢٠,٦٢ ٢٠,٨٨ ٢٠,٩٣ ٢٠,٩٨ ٢٠,٩٩ ٢١,١٥ ٢٢,١٦

٢٢,٢٤ ٢٣,١٦ ٢٣,٥٦ ٣٥,٧٨ ٣٦,٧٣

يلاحظ أن :

قيمة الربيع الأول	=	١٨,٢٧
قيمة الربيع الثالث	=	٢٢,١٦
قيمة الوسيط	=	٢٠,٨٨
أصغر مشاهدة	=	١٢,٨١
اكبر مشاهدة	=	٣٦,٧٣

(ويترك للقارىء التأكد من صحة هذه النتائج) .



شكل (٢)

ويظهر شكل الصندوق والشعيرات المناظر لهذه البيانات في شكل (٢) ، حيث يلاحظ أن الجانب الأيسر للصندوق يناظر ١٨,٢٧ (الربيع الأول) والجانب الأيمن يناظر ٢٢,١٦ (الربيع الثالث) ، بحيث يمثل عرض الصندوق ٤,١١ وهي قيمة المدى الربيعي . وتمتد الشعيرة اليسرى حتى قيمة أصغر مشاهدة (١٢,٨١) بينما تمتد الشعيرة اليمنى حتى قيمة اكبر مشاهدة (٣٦,٧٣) .

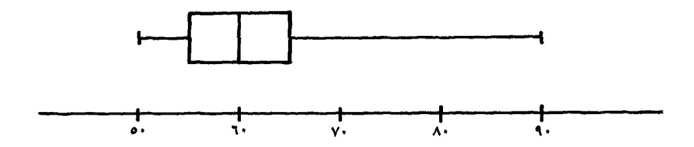
إذا كان توزيع المشاهدات في النصف الأوسط للبيانات متماثلاً فإن الخط الأفقي الممثل للوسيط يقع عند منتصف الصندوق . ويلاحظ في شكل (٢) أن خط الوسيط يقع إلى يمين منتصف الصندوق وأن الشعيرة اليمنى أطول من الشعيرة اليسرى . ويدل ذلك على وجود التواء لليمين في البيانات وعلى وجود بعض المشاهدات المتطرفة في هذا الاتجاه .

مثال (٤) : ارسم شكل الصندوق والشعيرات المناظر لبيانات مثال (٢) صفحة (٢٥٢) .

الحل : يلاحظ في هذا المثال أن :

قيمة الربيع الأول = ٥٧,٢١
قيمة الربيع الثالث = ٦٤,٥٦
قيمة الوسيط = ٦٠,٨٨
أصغر مشاهدة = ٥٠
اكبر مشاهدة = ٨٤

ويظهر شكل الصندوق والشعيرات المناظر في شكل (٣) .



شكل (٣) : شكل الصندوق والشعيرات

مثال (٥) : يمكن الاعتماد على شكل الصندوق والشعيرات للمقارنة بين نمط الاختلاف في مجموعات البيانات المختلفة . كمثال على ذلك ارسم شكل الصندوق والشعيرات المناظر لكل من المجموعة (أ) والمجموعة (جـ) من البيانات في شكل (١) .

الحل : يلاحظ في هذه البيانات أن :

للمجموعة (أ) : قيمة الربيع الأول = ٣٠
قيمة الربيع الثالث = ٦٠
قيمة الوسيط = ٤٥

اصغر مشاهدة = ٢٠

اكبر مشاهدة = ٧٠

للمجموعة (حـ) : قيمة الربيع الأول = ٤٣

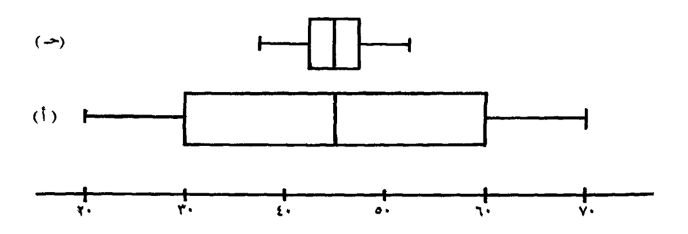
قيمة الربيع الثالث = ٤٧

قيمة الوسيط = ٤٥

اصغر مشاهدة = ٤٠

اكبر مشاهدة = ٥٠

وتـظهر أشكـال الصندوق والشعيـرات المناظـرة في شكل (٤) ، حيث يظهر بوضوح عدم تشابـه نمط الاختلاف في المجمـوعتين . ويترك للقـارىء وصف هذه الاختلافات كما تظهر في الشكل .



شكل (٤) : شكل الصندوق والشعيرات لبيانات المجموعة (أ) وبيانات المجموعة (حـ) في شكل (١) .

٣ ـ الانحراف المتوسط

تعتمد مقاييس التشتت السابقة الذكر مباشرة على جزء محدود فقط من المشاهدات المعطاة . فمثلاً يعتمد المدى على اكبر مشاهدة وأصغر مشاهدة

فقط ، على حين يعتمد المدى الربيعي على قيمة الربيع الأول وقيمة الربيع الثالث للمشاهدات . وعلى الرغم من أهمية هذه المقاييس في كثير من التطبيقات العملية ، إلا أن الحاجة تنشأ في معظم الأحيان إلى مقياس للتشتت يأخذ في الاعتبار اختلاف كل مشاهدة من المشاهدات عن المتوسط .

يعرف الانحراف المتوسط بأنه الوسط الحسابي للقيم المطلقة لانحرافات كل مشاهدة عن المتوسط. فمثلاً، إذا كانت الدرجات التي حصل عليها خمسة من الطلبة في أحد الامتحانات هي ٨، ٦، ٧، ٥، ٩ فإن الوسط الحسابي = $\frac{٣٥}{٥}$ = ٧ . إذا أريد انشاء مقياس لدرجة تشتت المشاهدات حول وسطها فإنه من المنطقي أن تحسب فروق أو انحرافات كل مشاهدة عن الوسط الحسابي . هذه الانحرافات هي : ٨ ـ ٧ = ١ ، ٦ ـ ٧ = ـ ١، ٧ ـ ٧ = صفر، ٥ ـ ٧ = ـ ٢ ، ٩ ـ ٧ = ٢ . حيث يلاحظ أن بعض هذه الانحرافات موجبة وبعضها الآخر سالبة ، وأن مجموع الانحرافات الموجبة يساوي دائماً مجموع قيم الانحرافات السالبة كما هو معروف عن خصائص الوسط الحسابي .

توضح القيمة المطلقة لكل انحراف (أي قيمة الانحراف بصرف النظر عن إشارتها) مدى اختلاف أو بعد المشاهدة المناظرة عن المتوسط . وعلى ذلك فإن الوسط الحسابي لهذه القيم المطلقة يمكن أن يؤخذ كمقياس لتشتت المشاهدات حول متوسطها . وتكتب القيمة المطلقة لكمية ما بوضع هذه الكمية داخل خطين رأسيين ، فمثلاً |ـ٣| = ٣ . ويكون الانحراف المتوسط للدرجات هو :

$$\frac{|١| + |-١| + |\text{صفر}| + |-٢| + |٢|}{٥} = \frac{١ + ١ + \text{صفر} + ٢ + ٢}{٥} = ١,٢$$

وتكون قيمة هذا المقياس صغيرة كلما كانت المشاهدات مركزة حول وسطها الحسابي ، والعكس تكون قيمته كبيرة اذا كانت المشاهدات مشتتة بشكل واضح حول هذا المتوسط .

ويمكن التعبير عن الانحراف المتوسط بشكل رياضي كما يلي . اذا كانت هناك مشاهدات عددها $ن$ هي $س_١$، $س_٢$، . . . $س_ن$ وكان الوسط الحسابي لهذه المشاهدات هو $\bar{س}$ ، فإن المسافة بين أي مشاهدة $س_ر$ والوسط الحسابي $\bar{س}$ تساوي $|س_ر - \bar{س}|$ ، أي القيمة الموجبة للفرق بين $س_ر$، $\bar{س}$ وذلك لقيم $ر$ = ١، ٢، ...، $ن$. ويكون الانحراف المتوسط هو الوسط الحسابي لهذه المسافات أي الوسط الحسابي للقيم الموجبة لانحراف المشاهدات عن وسطها . ويكتب ذلك كما يلي :

$$\text{الانحراف المتوسط} = \frac{\sum |س - \bar{س}|}{ن}$$

ويمكن حساب الانحراف المتوسط اذا كانت البيانات معطاة في شكل جدول تكراري . ويتضح كيفية ذلك في المثالين التاليين .

مثال (٦) : يعطي الجدول التالي التوزيع التكراري لعدد غرف المسكن في عينة من ١٠ مساكن أسرية ، والمطلوب حساب قيمة الانحراف المتوسط ، وتفسير معناه .

عدد الغرف	٣	٤	٥	٦	المجموع
عدد المساكن	١	٥	٢	٢	١٠

الحل : لما كان المتغير محل الدراسة (عدد غرف المسكن) متغيراً متقطعاً ، فإنه يمكن استخدام القواعد المعتادة لحساب الوسط الحسابي من أجل حساب الانحراف المتوسط . إذا كان هناك مشاهدات عددها $ن$ هي $س_١$ ، $س_٢$ . . . ، $س_ن$ وكان هناك م من القيم المتميزة للمتغير يرمز لها بالرمز $ف_١$ ، $ف_٢$ ، . . . ، $ف_م$ ، وكانت التكرارات المناظرة لهذه القيم هي $ك_١$ ، $ك_٢$ ، . . . ، $ك_م$ على الترتيب فإن الصيغة الرياضية لحساب الانحراف المتوسط في هذه الحالة هي :

$$\text{الانحراف المتوسط} = \frac{\text{مجـ ك} |\text{ف} - \bar{\text{س}}|}{\text{مجـ ك}}$$

أي أن خطوات الحساب تتمثل فيما يلي :

(١) حساب قيمة الوسط الحسابي $\bar{\text{س}}$

(٢) حساب القيمة المطلقة للفرق بين كل قيمة للمتغير في الجدول وبين $\bar{\text{س}}$ ، أي حساب قيم $|\text{ف} - \bar{\text{س}}|$

(٣) ضرب كل قيمة مطلّقة للانحرافات في التكرار المناظر لها ونجمع حواصل الضرب لنحصل على $\text{مجـ ك} |\text{ف} - \bar{\text{س}}|$

(٤) نطبق الصيغة الرياضية للانحراف المتوسط .

وتتضح خطوات العمل هذه في الجدول الحسابي التالي .

عدد الغرف (ف)	عدد المساكن (ك)	ك ف	$|\text{ف} - \bar{\text{س}}|$	$\text{ك} |\text{ف} - \bar{\text{س}}|$
٣	١	٣	١,٥ = |٣ - ٤,٥|	١ × ١,٥ = ١,٥
٤	٥	٢٠	,٥ = |٤ - ٤,٥|	٥ × ٢,٥ = ١٢,٥
٥	٢	١٠	٠,٥ = |٥ - ٤,٥|	٢ × ٠,٥ = ١,٠
٦	٢	١٢	١,٥ = |٦ - ٤,٥|	٢ × ١,٥ = ٣,٠
المجموع	١٠	٤٥		١٨,٠٠

$$\text{الوسط الحسابي } \bar{\text{س}} = \frac{٤٥}{١٠} = ٤,٥ \text{ غرفه}$$

$$\text{الانحراف المتوسط} = \frac{١٨}{١٠} = ١,٨ \text{ غرفة}$$

ويعني ذلك أن الوسط الحسابي لعدد غرف المسكن = ٤,٥ غرفة وأن

المساكن تختلف حول هذا المتوسط بحيث أن متوسط البعد بين عدد غرف كل مسكن والمتوسط يساوي ١,٨ غرفة .

مثال (٧): يعطي الجدول الآتي التوزيع التكراري للعمر في عينة من ٢٠٠ موظف ، والمطلوب حساب قيمة الانحراف المتوسط لهذه الأعمار .

فئات العمر بالسنوات	٢٠ ـ ٢٩	٣٠ ـ ٣٩	٤٠ ـ ٤٩	٥٠ ـ ٥٩	٦٠ ـ ٦٩	المجموع
عدد الموظفين	٣٠	٧٠	٥٠	٣٠	٢٠	٢٠٠

الحل :

يلاحظ أن المتغير محل الدراسة (عمر الموظف) متغير متصل . وفي هذه الحالة يحسب الانحراف المتوسط بنفس الطريقة التي ذكرت في مثال (٦) بعد أخذ قيمة مركز كل فئة للتعبير عن القيمة المتوسطة لها والتي تستخدم كقيم ف في صيغة الحساب . وتظهر خطوات الحساب في الجدول التالي :

فئات العمر بالسنوات	عدد الموظفين (ك)	مركز الفئة (ف)	ك ف	$\lvert ف - \bar{س} \rvert$	ك $\lvert ف - \bar{س} \rvert$
٢٠ ـ ٢٩	٣٠	٢٤,٥	٧٣٥	$\lvert ٢٤,٥ - ٤١,٥ \rvert = ١٧$	٣٠ × ١٧ = ٥١٠
٣٠ ـ ٣٩	٧٠	٣٤,٥	٢٤١٥	$\lvert ٣٤,٥ - ٤١,٥ \rvert = ٧$	٧٠ × ٧ = ٤٩٠
٤٠ ـ ٤٩	٥٠	٤٤,٥	٢٢٢٥	$\lvert ٤٤,٥ - ٤١,٥ \rvert = ٣$	٥٠ × ٣ = ١٥٠
٥٠ ـ ٥٩	٣٠	٥٤,٥	١٦٣٥	$\lvert ٥٤,٥ - ٤١,٥ \rvert = ١٣$	٣٠ × ١٣ = ٣٩٠
٦٠ ـ ٦٩	٢٠	٦٤,٥	١٢٩٠	$\lvert ٦٤,٥ - ٤١,٥ \rvert = ٢٣$	٢٠ × ٢٣ = ٤٦٠
المجموع	٢٠٠		٨٣٠٠		٢٠٠٠

$$\text{الوسط الحسابي } \bar{س} = \frac{٨٣٠٠}{٢٠٠} = ٤١,٥ \text{ سنة}$$

$$\text{الانحراف المتوسط} = \frac{\text{مجـ ك} \lvert ف - \bar{س} \rvert}{\text{مجـ ك}} = \frac{٢٠٠٠}{٢٠٠} = ١٠ \text{ سنوات .}$$

معنى ذلك أن الوسط الحسابي لعمر الموظف = ٤١,٥ سنة ، كما أن كل عمر يختلف في المتوسط عن هذا المركز بمقدار عشر سنوات ، مما يدل على وجود اختلافات كبيرة بين هذه الأعمار .

ويمكن أخذ إنحرافات المشاهدات عن أي مقياس للنزعة المركزية ، اذ قد تحسب الانحرافات عن الوسيط مثلاً بدلاً من حسابها عن الوسط الحسابي . وتجدر الإشارة إلى أن مجموع المسافات بين المشاهدات والوسيط يكون دائماً أقل من مجموع المسافات بين المشاهدات والوسط الحسابي (أو بين المشاهدات وأي قيمة أخرى ثابتة) .

ويعتبر الانحراف المتوسط مقياساً وصفياً جيداً للتشتت ، إلا أنه يعاني من أوجه قصور أهمها صعوبة اخضاعه للعمليات الرياضية . وينعكس ذلك بشكل خاص في عدم شيوع استخدام هذا المقياس لأغراض الاستنتاج الاحصائي .

٤ ـ التباين والانحراف المعياري

يقيس التباين درجة التشتت في المشاهدات حول وسطها الحسابي . ويعتمد التباين في هذا الصدد ، شأنه في ذلك شأن الانحراف المتوسط ، على انحرافات المشاهدات عن الوسط .

إذا كان هناك مشاهدات عددها $ن$ هي $س_١$ ، $س_٢$ ، . . . ، $س_ن$ وكان وسطها الحسابي هو $\bar{س}$ فإن انحرافات المشاهدات عن الوسط هي $س_١ - \bar{س}$ ، $س_٢ - \bar{س}$ ، . . . ، $س_ن - \bar{س}$. ومن المنطقي الاعتماد على هذه الانحرافات لدراسة درجة التشتت في المشاهدات ، وذلك من خلال وضعها في مقياس مناسب . وقد سبقت الإشارة إلى أن بعض قيم هذه الانحرافات تكون موجبة وبعضها الآخر تكون سالبة بحيث يكون مجموع القيم الموجبة يساوي دائماً مجموع القيم السالبة ويكون مجـ $(س - \bar{س})$ = صفر . ويجب التنويه الى أن درجة التشتت في البيانات تنعكس في أحجام هذه الانحرافات وليس في اشاراتها .

يتطلب انشاء مقياس مناسب لدرجة التشتت التخلص من إشارات انحرافات المشاهدات عن وسطها الحسابي . ويمكن تحقيق ذلك بأحد أسلوبين :

أ ـ التغاضي عن الإشارة والاعتماد على القيم المطلقة للانحرافات . ويمثل هذا الأسلوب أساس حساب الانحراف المتوسط الذي ينتج بأخذ متوسط هذه القيم المطلقة .

ب ـ تربيع الانحرافات للتخلص من إشاراتها . ويمثل ذلك أساس حساب كل من التباين والانحراف المعياري .

إذا كانت المشاهدات تمثل بيانات مجتمع حجمه ن فإن التباين يكون متوسط مربعات انحرافات المشاهدات عن وسطها الحسابي . فإذا كانت المشاهدات هي $س_١$ ، $س_٢$ ، . . . ، $س_ن$ ووسطها الحسابي هو μ (لأنه متوسط مجتمع) فإن التباين ويرمز له بالرمز $\sigma^٢$ يعطى بالعلاقة

$$\sigma^٢ = \frac{مجـ (س - \mu)^٢}{ن}$$

ويكون الانحراف المعياري هو الجذر التربيعي الموجب للتباين ويرمز له بالرمز σ أي أن :

$$\sigma = \sqrt{\frac{مجـ (س - \mu)^٢}{ن}}$$

فمثلاً إذا كانت أوزان مجتمع مكون من خمسة أطفال هي ٨ ، ٦ ، ٧ ، ٥ ، ٩ كيلوجرام فإن الوسط الحسابي :

$$\mu = \frac{٨ + ٦ + ٧ + ٥ + ٩}{٥} = \frac{٣٥}{٥} = ٧ \text{ كيلوجرام .}$$

وتكون قيمة التباين σ^2 هي :

$$\sigma^2 = \frac{(٨ - ٧)^2 + (٦ - ٧)^2 + (٧ - ٧)^2 + (٥ - ٧)^2 + (٩ - ٧)^2}{٥}$$

$$= \frac{١٠}{٥} = ٢ \text{ (كيلو جرام)}^2 .$$

وتكون قيمة الانحراف المعياري σ هي :

$$\sigma = \sqrt{٢} = ١,٤١ \text{ كيلوجرام} .$$

ويلاحظ أن وحدات قياس الانحراف المعياري هي نفس وحدات قياس المتغير محل الدراسة . كذلك فإن وحدات قياس التباين تكون وحدات قياس المتغير مربعة .

أما إذا كانت المشاهدات تمثل بيانات عينة حجمها $ن$ فإن التبـاين في هذه الحالة ويرمز له بالرمز $ع^2$ ينتج بالقسمة على ($ن$ - ١) بدلاً من $ن$. أي أن :

$$\text{تباين العينة } ع^2 = \frac{\text{مجـ} (س - \bar{س})^2}{ن - ١}$$

ويكون الانحراف المعياري في العينة ع هو الجذر التـربيعي الموجب لهذا التباين . فمثلاً إذا كان عدد أطفال الأسرة في عينة ٧ أسر هي ١ ، ٤ ، ٦ ، ١ ، ٦ ، ٢ ، ١ فإن الوسط الحسابي : $\bar{س} = \frac{٢١}{٧} = ٣$ أطفال .

وتكون قيمة التباين في العينة $ع^2$ هي :

$$ع^2 = \frac{(١ - ٣)^2 + (٤ - ٣)^2 + (٦ - ٣)^2 + (١ - ٣)^2 + (٦ - ٣)^2 + (٢ - ٣)^2 + (١ - ٣)^2}{٧ - ١}$$

$$= \frac{٣٢}{٦} = ٥,٣٣ \text{ (طفل)}^٢$$

وتكون قيمة الانحراف المعياري في العينة ع هي :

$$ع = \sqrt{٥,٣٣} = ٢,٣١ \text{ طفل .}$$

ويرجع استخدام (ﻥ - ١) بدلاً من ﻥ عند حساب تباين العينة $ع^٢$ إلى أن إحصاءات العينة تستخدم أساساً لتقدير معالم المجتمع . وتشير النظرية الاحصائية إلى أن الاحصاء $ع^٢$ يكون أكثر جودة كمقدر للمعلمة $\sigma^٢$ إذا قسم على (ﻥ - ١) بدلاً من ﻥ ، وخاصة إذا كان حجم العينة صغيراً .

ويعتبر الانحراف المعياري أكثر المقاييس المستخدمة للتشتت وذلك على الرغم من صعوبة تفسير معناه إذا ما قورن بالمدى أو الانحراف المتوسط . ويرجع ذلك إلى أن صيغته الرياضية تناسب أغراض الاستنتاج الإحصائي كما يمكن تفسيره بأنه مقياس لمتوسط بعد القيم عن وسطها الحسابي . ويستخدم الانحراف المعياري بدلاً من التباين لقياس التشتت لأن الانحراف المعياري يقاس بنفس وحدات قياس المتغير محل الدراسة ، على عكس التباين الذي يقاس بمربع هذه الوحدات .

لما كان الانحراف المعياري يعتمد على تربيع انحرافات المشاهدات عن وسطها الحسابي ، فإن ذلك يعني تأثره بالقيم الشاذة أو المتطرفة والتي يناظرها عادة قيماً كبيرة لهذه الانحرافات . لتوضيح ذلك، فيما يلي بيانات عن فترة الحضانة للتسمم بالطعام (وهي الفترة التي تنقضي بين أكل طعام فاسد وظهور أعراض التسمم) لمجموعة من ٢١ شخصاً أصيبوا بالتسمم في إحدى المدن (الفترة مقاسة بالساعات) . ويراد دراسة الخصائص العامة لفترة حضانة هذا المرض . وبصفة خاصة يراد التعرف على متوسط طول فترة الحضانة وعلى مقياس لدرجة التشتت في هذه البيانات .

٣٦	٤٨	١٤
٤٣	٤٥	١٩
٢٩	٤٨	٢٠
١٩	٢١	٢٠
٤٣	٧٨	٣٢
٨٥	١٧	٣٤
٩١	٢٠	٣٦

يوضح الجدول التالي حساب الوسط الحسابي $\bar{س}$ والانحراف المعياري ع لهذه البيانات ، حيث يلاحظ أن : $\bar{س}$ = ٣٨ ساعة وأن ع = ٢٢,٤ ساعة . يلاحظ أن قيمة الانحراف المعياري كبيرة بالمقارنة بالوسط الحسابي . في مثل هذه الحالات يجب فحص البيانات للتعرف على سبب ذلك . ويرجع كبر قيمة ع في معظم الحالات إلى وجود قيم شاذة في البيانات . فمثلاً يلاحظ من الحسابات أن المشاهدات (٧٨ ، ٨٥ ، ٩١) كبيرة بالنسبة لباقي المشاهدات وأن نصيب هذه المشاهدات من مجموع مربعات انحرافات القيم عن وسطها الحسابي يزيد عن ٦٥٪ من هذا المجموع :

$$\left(\frac{\text{١٦٠٠} + \text{٢٢٠٩} + \text{٢٨٠٩}}{\text{١٠٠٥٨}} = \text{٠,٦٥٨} \right) .$$

أي أن هذه القيم الكبيرة تمثل السبب الرئيسي لكبر قيمة الانحراف المعياري .

س	س - $\bar{س}$	(س - $\bar{س}$)^{٢}
١٤	- ٢٤	٥٧٦
١٩	- ١٩	٣٦١
٢٠	- ١٨	٣٢٤
٢٠	- ١٨	٣٢٤
٣٢	- ٦	٣٦
٣٤	- ٤	١٦
٣٦	- ٢	٤
٤٨	+ ١٠	١٠٠
٤٥	+ ٧	٤٩
٤٨	+ ١٠	١٠٠
٢١	- ١٧	٢٨٩
٧٨	+ ٤٠	١٦٠٠
١٧	- ٢١	٤٤١
٢٠	- ١٨	٣٢٤
٣٦	- ٢	٤
٤٣	+ ٥	٢٥
٢٩	- ٩	٨١
١٩	- ١٩	٣٦١
٤٣	+ ٥	٢٥
٨٥	+ ٤٧	٢٢٠٩
٩١	+ ٥٣	٢٨٠٩
٧٩٨		١٠٠٥٨

($\bar{س} = \frac{٧٩٨}{٢١} = ٣٨$ ، $ع = \sqrt{\frac{١٠٠٥٨}{٢٠}} = \sqrt{٥٠٢,٩} = ٢٢,٤$)

ويلاحظ كذلك أنه طالما أن هذه القيم الثلاث تقع جميعها إلى يمين الوسط الحسابي ، فإن ذلك يعني وجود التواء لليمين في المشاهدات . وينصح في مثل هذه المواقف بالاعتماد على مقاييس أخرى للتشتت مثل الانحراف المتوسط أو المدى الربيعي . وتجدر الإشارة إلى أن الانحراف المتوسط يكون أقل حساسية من الانحراف المعياري للقيم الشاذة في البيانات . وسوف نرى فيما بعد أن استخدام الوسط الحسابي والانحراف المعياري معاً لوصف مجموعة من المشاهدات يكون مقبولاً اذا لم تكن هناك قيم متطرفة في البيانات وإذا كان شكل التوزيع المناظر يقترب من شكل المنحنى الطبيعي .

يلاحظ أن حساب كل من التباين والانحراف المعياري يتطلب حساب الفرق بين كل مشاهدة والوسط الحسابي . وقد يؤدي ذلك عملياً إلى صعوبات في الحساب ، خاصة اذا كان الوسط الحسابي يحتوي على عدد كبير من الأرقام المعنوية ، كما قد يترتب عليه انخفاض في دقة الحسابات نتيجة عمليات التقريب المستخدمة . لذلك ، يفضل عادة حساب التباين والانحراف المعياري مباشرة دون حساب الانحرافات عن المتوسط بشكل صريح . ويتم ذلك بملاحظة أن :

$$\text{مجـ}(\text{س} - \bar{\text{س}})^{\text{٢}} = \text{مجـ}(\text{س}^{\text{٢}} - \text{٢}\,\bar{\text{س}}\,\text{س} + \bar{\text{س}}^{\text{٢}})$$
$$= \text{مجـ س}^{\text{٢}} - \text{مجـ ٢ س}\,\bar{\text{س}} + \text{مجـ}\,\bar{\text{س}}^{\text{٢}}$$
$$= \text{مجـ س}^{\text{٢}} - \text{٢}\,\bar{\text{س}}\,\text{مجـ س} + \text{ن}\,\bar{\text{س}}^{\text{٢}}$$
$$= \text{مجـ س}^{\text{٢}} - \text{٢}\,\bar{\text{س}}\,(\text{ن س}) + \text{ن س}^{\text{٢}}$$
$$= \text{مجـ س}^{\text{٢}} - \text{٢ن}\,\bar{\text{س}}^{\text{٢}} + \text{ن}\,\bar{\text{س}}^{\text{٢}}$$
$$= \text{مجـ س}^{\text{٢}} - \text{ن}\,\bar{\text{س}}^{\text{٢}} .$$

وبالتعويض عن قيمة :

$$\bar{\text{س}} = \frac{\text{مجـ س}}{\text{ن}}$$ ينتج أن :

$$\text{مجـ}(س - \bar{س})^2 = \text{مجـ}س^2 - \frac{(\text{مجـ}س)^2}{ن}$$

وعلى ذلك تكون الصيغة الحسابية للتباين هي :

$$ع^2 = \frac{1}{ن - 1}\left[\text{مجـ}س^2 - \frac{(\text{مجـ}س)^2}{ن}\right]$$

مثال ٨ : احسب قيمة التباين للمشاهدات التالية التي تمثل عدد غرف المسكن في عينة من ٧ مساكن :

١ ، ٤ ، ٦ ، ١ ، ٦ ، ٢ ، ١

الحل : لإيجاد قيمة $ع^2$ ، يلزم حساب مجـ س ، مجـ $س^2$ وتظهر الحسابات فيما يلي :

س	$س^2$
١	١
٤	١٦
٦	٣٦
١	١
٦	٣٦
٢	٤
١	١
٢١	٩٥

ويكون التباين :

$$ع^2 = \frac{1}{ن - 1}\left[\text{مجـ}س^2 - \frac{(\text{مجـ}س)^2}{ن}\right]$$

$$= \frac{١}{٧ - ١} \left[٩٥ - \frac{(٢١)^٢}{٧} \right]$$

$$= \frac{١}{٦} \left[٩٥ - \frac{٤٤١}{٧} \right] = \frac{٣٢}{٦} = ٥,٣٢ \text{ غرفة .}$$

ويمكن أيضاً حساب التباين والانحراف المعياري إذا كانت البيانات معطاه في جدول توزيع تكراري . ويعتمد في هذه الحالة على القواعد العامة لحساب الوسط الحسابي ويكون التباين $ع^٢$ معطى بالعلاقة .

$$ع^٢ = \frac{مجـ ك (ف - \bar{س})^٢}{مجـ ك - ١}$$

حيث قيم ك ، ف ترمز الى التكرارات وقيم المتغير (أو مراكز الفئات) على الترتيب كما سبق القول عند الحديث عن كيفية حساب الوسط الحسابي والانحراف المتوسط .

كذلك يمكن إعادة كتابة هذه الصيغة على الشكل :

$$ع^٢ = \frac{١}{مجـ ك - ١} \left[مجـ ك ف^٢ - \frac{(مجـ ك ف)^٢}{مجـ ك} \right]$$

ويكون الانحراف المعياري ع هو الجذر التربيعي الموجب لهذا التباين . ويوضح المثالين التاليين كيفية حساب التباين والانحراف المعياري من جدول تكراري .

مثال(٩) : يعطي الجدول التالي التوزيع التكراري لعدد الإجابات الصحيحة لكل طالب في أحد الامتحانات وذلك في عينة من ٥٠ طالباً . والمطلوب حساب قيمة الانحراف المعياري لعدد الدرجات .

عدد الاجابات الصحيحة	٤	٥	٦	٧	٨	٩	١٠	المجموع
عدد الطلبة	٥	٧	٨	١٠	٨	٧	٥	٥٠

الحل : يلزم لحساب قيمة الانحراف المعياري ايجاد قيم المجاميع مجـ ك ف ، مجـ ك ف^{٢} . ويظهر ذلك في الجدول التالي :

عدد الاجابات ف	عدد الطلبة ك	ك ف	ك ف^{٢}
٤	٥	٢٠	٨٠
٥	٧	٣٥	١٧٥
٦	٨	٤٨	٢٨٨
٧	١٠	٧٠	٤٩٠
٨	٨	٦٤	٥١٢
٩	٧	٦٣	٥٦٧
١٠	٥	٥٠	٥٠٠
	٥٠	٣٥٠	٢٦١٢

ويكون التباين : $ع^{٢} = \frac{١}{مجـ ك - ١} \left[مجـ ك ف^{٢} - \frac{(مجـ ك ف)^{٢}}{مجـ ك} \right]$

$$= \frac{١}{٥٠ - ١} \left[٢٦١٢ - \frac{(٣٥٠)^{٢}}{٥٠} \right]$$

$$= \frac{١}{٤٩} \left[٢٦١٢ - ٢٤٥٠ \right] = ٣,٣٠٦$$

ويكـون الانحراف المعيـاري $ع = \sqrt{ع^{٢}} = \sqrt{٣,٣٠٦} = ١,٨٢$

إجابة صحيحة . ويلاحظ أيضاً أن الحسابات السابقة تمكن من إيجاد قيمة الوسط الحسابي مباشرة ، حيث نجد أن :

$$\bar{س} = \frac{مجـ ك ف}{مجـ ك} = \frac{٣٥٠}{٥٠} = ٧$$ إجابات صحيحة .

مثال (١٠) : يعطي الجدول التالي مستوى الكولسترول في الدم (بالميليجرام لكل ١٠٠ ميلمتر) لعينة من ٢٠٩ شخصاً . والمطلوب حساب قيمة الوسط الحسابي والانحراف المعياري .

مستوى الكولسترول	١٢٥ ـ ١٤٩	١٥٠ ـ ١٧٤	١٧٥ ـ ١٩٩	٢٠٠ ـ ٢٢٤
عدد الأشخاص	٤	١٣	٣٠	٤٢
مستوى الكولسترول	٢٢٥ ـ ٢٤٩	٢٥٠ ـ ٢٧٤	٢٧٥ ـ ٢٩٩	٣٠٠ ـ ٣٢٤
عدد الأشخاص	٥١	٣٤	٢٣	٨
مستوى الكولسترول	٣٢٥ ـ ٣٤٩	٣٥٠ ـ ٣٧٤		المجموع
عدد الأشخاص	٣	١		٢٠٩

الحل : نبدأ بحساب مجـ ك ف ، مجـ ك $ف^{٢}$ كما يظهر في الجدول التالي :

مستوى الكولسترول	عدد الأشخاص ك	مركز الفئة ف	ك ف	ك $ف^{٢}$
١٢٥ ـ ١٤٩	٤	١٣٧	٥٤٨	٧٥٠٧٦
١٥٠ ـ ١٧٤	١٣	١٦٢	٢١٠٦	٣٤١١٧٢
١٧٥ ـ ١٩٩	٣٠	١٨٧	٥٦١٠	١٠٤٩٠٧٠
٢٠٠ ـ ٢٢٤	٤٢	٢١٢	٨٩٠٤	١٨٨٧٦٤٨
٢٢٥ ـ ٢٤٩	٥١	٢٣٧	١٢٠٨٧	٢٨٦٤٦١٩
٢٥٠ ـ ٢٧٤	٣٤	٢٦٢	٨٩٠٨	٢٣٣٣٨٩٦
٢٧٥ ـ ٢٩٩	٢٣	٢٨٧	٦٦٠١	١٨٩٤٤٨٧
٣٠٠ ـ ٣٢٤	٨	٣١٢	٢٤٩٦	٧٧٨٧٥٢
٣٢٥ ـ ٣٤٩	٣	٣٣٧	١٠١١	٣٤٠٧٠٧
٣٥٠ ـ ٣٧٤	١	٣٦٢	٣٦٢	١٣١٠٤٤
المجموع	٢٠٩		٤٨٦٣٣	١١٦٩٦٤٧١

ويكون الوسط الحسابي :

$$\bar{س} = \frac{مجـ ك ف}{مجـ ك} = \frac{٤٨٦٣٣}{٢٠٩} = ٢٣٢,٦٩ \text{ ميلجرام/١٠٠ ميليمتر}$$

كما أن التباين $ع^2$

$$= \frac{١}{مجـ ك - ١}\left[مجـ ك ف^2 - \frac{(مجـ ك ف)^2}{مجـ ك}\right]$$

$$= \frac{١}{٢٠٩ - ١}\left[١١٦٩٦٤٧١ - \frac{(٤٨٦٣٣)^2}{٢٠٩}\right]$$

$$= \frac{١}{٢٠٨}\left[١١٦٩٦٤٧١ - ١١٣١٦٢٢٩\right]$$

$$= \frac{٣٨٠٢٤٢}{٢٠٨} = ١٨٢٨,١$$

وتكون قيمة الانحراف المعياري $ع = \sqrt{١٨٢٨,١} = ٤٢,٨$ ميلجرام/١٠٠ ميليمتر .

٥ ـ إختيار المقياس الملائم للتشتت

سبقت الإشارة إلى أن مقاييس التشتت تستخدم في غرضين أساسيين هما :

أ ـ تقويم مدى كفاءة المتوسط في وصف موضع التوزيع . وقد جرت العادة ، في هذا الصدد ، على اعطاء قيمة مقياس التشتت جنباً إلى جنب مع مقياس الموضع . فمثلاً إذا كانت قيمة وسيط الدخل للأسرة في بلد ما = ٧٢٠٠ درهم فإن هذا الوسيط يكون اكثر تمثيلاً للنزعة المركزية في البيانات اذا كان المدى الربيعي المناظر = ٢٠٠٠ درهم بالمقارنة بالحالة التي يكون فيها المدى الربيعي مساوياً ٦٠٠٠ درهم .

كذلك فإن الانحراف المعياري يدل على مدى كفاءة الوسط الحسابي في تمثيل مركز البيانات ، بحيث يكون الوسط الحسابي اكثر جودة كلما كانت قيمة الانحراف المعياري صغيرة .

ب - إجراء المقارنة بين نمط الاختلاف في توزيعين أو أكثر . ذلك أن هذه المقارنة تتم عادة بمقارنة كل من الموضع والتشتت في التوزيعين . وقد تكون هناك حالات تختلف فيها التوزيعات حسب الموضع فقط أو حسب التشتت فقط أو حسب كلاهما .

وتجدر الإشارة إلى أن اختيار مقياس التشتت المناسب يعتمد على ثلاث عوامل رئيسية هي :

أ - مفهوم مقياس التشتت المطلوب في الدراسة . وقد أشرنا سابقاً إلى أن المدى يستخدم في تطبيقات مراقبة جودة الانتاج وبعض التطبيقات الصحية ، كذلك يستخدم المدى الربيعي في الحالات التي يكون فيها دراسة الترتيب بين المشاهدات مفيداً مثل دراسة الدرجات التي يحصل عليها الطلبة أو الدخول التي تحصل عليها الأسر . أما الانحراف المتوسط والانحراف المعياري فهي مقاييس تحسب اذا كان المطلوب مقياساً لاختلافات المشاهدات حول وسطها .

ب - نوع البيانات المتاحة . لا ينصح باستخدام الانحراف المعياري اذا كان عدد المشاهدات قليلاً أو إذا كانت هناك مشاهدات متطرفة أو شاذة . كذلك يفضل تجنب حساب الانحراف المتوسط إذا كانت درجة الالتواء مرتفعة في البيانات. كما يكون المدى مقياساً غير مفيد اذا كانت هناك اختلافات واسعة بين المشاهدات .

جـ - خصائص المقاييس المستخدمة . وفي هذا الصدد ، يعتبر الانحراف المعياري اكثر مقاييس التشتت المستخدمة ، لما لهذا المقياس من خصائص مرغوبة . وتشمل هذه الخصائص اعتماده على جميع

المشاهدات المعطاة وسهولة التعامل معه جبرياً مما يسهل استخداماته لأغراض الاستنتاج الاحصائي . ولذلك ينصح باستخدام الانحراف المعياري ، اذا كان مقياس التشتت سوف يستخدم في خطوات تحليلية تالية . ويستخدم الانحراف المتوسط اذا أريد اعطاء اوزان متساوية لجميع انحرافات المشاهدات عن متوسطها . بينما يستخدم الوسيط والربيعين في الحالات التي يكون فيها الالتواء واضحاً في البيانات .

٦ ـ وصف وتلخيص التوزيع التكراري ـ قاعدة تشيبيتشيف

تهتم الأقسام الباقية في هذا الباب بكيفية استخدام مقياس النزعة المركزية مع مقياس التشتت للوصول إلى وصف موجز لتوزيع المشاهدات . وتعتمد الأساليب المستخدمة على ما هو معروف من أن درجة تركز المشاهدات حول وسطها تكون مرتفعة كلما كانت قيمة مقياس التشتت صغيرة . وتعتبر قاعدة تشيبيتشيف إحدى القواعد التي توضح ذلك بشكل رياضي بسيط .

إذا كانت البيانات المتاحة عن توزيع ما تشمل وسطه الحسابي وانحرافه المعياري فقط، فإن قاعدة تشيبيتشيف يمكن أن تستخدم لإعطاء معلومات عن نسبة المشاهدات التي تقع داخل بعد معين من الوسط الحسابي ، وذلك بافتراض أن هذا البعد يمكن كتابته كمضاعف للانحراف المعياري . وتنص هذه القاعدة بالتحديد أنه كان هناك مقدار ثابت ل قيمته تساوي على الأقل واحد فإن نسبة المشاهدات التي تقع داخل بعد يساوي ل ع من الوسط الحسابي تكون على الأقل مساوية ١ ـ $\frac{١}{ل^٢}$. وتكون هذه القاعدة صحيحة سواء كانت المشاهدات معطاة لمجتمع أو لعينة . ويوضح الجدول الآتي نتائج تطبيق هذه القاعدة لقيم مختلفة من قيم ل .

قيمة ل	نسبة المشاهدات التي تقع داخل الفترة $\bar{س}$ ـ ل ع ، $\bar{س}$ + ل ع
٢	على الأقل ١ - $\frac{١}{٤}$ أي على الأقل ٠,٧٥
٣	على الأقل ١ - $\frac{١}{٩}$ أي على الأقل ٠,٨٩
٤	على الأقل ١ - $\frac{١}{١٦}$ أي على الأقل ٠,٩٤
٤,٤٧٢	على الأقل ١ - $\frac{١}{٢٠}$ أي على الأقل ٠,٩٥
٥	على الأقل ١ - $\frac{١}{٢٥}$ أي على الأقل ٠,٩٦
١٠	على الأقل ١ - $\frac{١}{١٠٠}$ أي على الأقل ٠,٩٩

مثال (١٠) : إذا كان معلوماً أن الوسط الحسابي لوزن محتويات العلبة من الخضروات المحفوظة التي يعبئها أحد المصانع = ١٦,٠٠ أوقية وأن الانحراف المعياري لهذه الأوزان =٠,٠٢ أوقية ، استخدم قاعدة تشيبيتشيف للحصول على حد أدنى لنسبة العلب في انتاج هذا المصنع التي تتراوح أوزانها بين ١٥,٩٥أوقية ، ١٦,٠٥ أوقية .

الحل : اذ أريد وضع الفترة (١٥,٩٥ ، ١٦,٠٥) على الشكل ($\bar{س}$ - ل ع ، س + ل ع) فإن ذلك يعني أن :

١٥,٩٥ = ١٦,٠٠ - ٠,٠٢ ل

وأن ١٦,٠٥ = ١٦,٠٠ + ٠,٠٢ ل

ويستنتج من ذلك أن ٠,٠٢ ل = ٠,٠٥

$$\text{أو ل} = \frac{٠,٠٥}{٠,٠٢} = ٢,٥$$

وعلى ذلك فإن نسبة المشاهدات التي تقع بين $\bar{س}$ - ٢,٥ع ، $\bar{س}$ + ٢,٥ع تكون مساوية ١ - $\frac{١}{(٢,٥)^{٢}}$ على الأقل ، أي ١ - ٠,١٦ = ٠,٨٤

على الأقل . ويعني ذلك أن ١٦٪ على الأكثر من هذه العلب يكـون وزنها خارج الفترة (١٥,٩٥ ، ١٦,٠٥) .

وتستخدم قاعدة تشيبيتشيف مع جميع التوزيعات على السواء ، وذلـك دون اعتبار لشكل التوزيع أو سماته العامة . وينعكس ذلك في كون التقديرات التي تنشأ من تطبيق هذه القاعـدة تقديـرات محافـظة . فمثلاً ، يؤدي تـطبيق القاعدة إلى القول بأن ٧٥٪ على الأقل من المشاهدات تقع على بعد لا يتعدى انحرافين معياريين من الوسط الحسابي ، بينما يلاحظ في كثير من الحالات أن النسبة الفعلية في المشاهدات تزيد كثيراً عن ٠,٧٥ ويتضح ذلك من المثال التالي .

مثال (١١) : فيما يلي شكل الأغصان والأوراق لقيم مقياس الذكاء في

٧	١
٦	٩، ٧، ٦، ٥، ٢
٨	٨، ٦، ٦، ٥، ٥، ٥، ٤، ٢، ١، ٠، ٠، ٠، ٠،
٩	٩، ٨، ٨، ٨، ٧، ٧، ٦، ٦، ٦، ٦، ٤، ٤، ٣، ٣، ٣، ٢، ١، ١، ٠، ٠، ٠، ٠
١٠	،، ٩، ٩، ٩، ٩، ٨، ٨، ٧، ٧، ٧، ٧، ٦، ٦، ٦، ٥، ٣، ٣، ٣، ٢، ٢، ٢، ٢، ٢، ١، ١، ٠، ٠، ٠
١١	٩، ٩، ٨، ٧، ٧، ٤، ٤، ٤، ٤، ٤، ٤، ٣، ٣، ٣، ٣، ٢، ٢، ١، ١، ٠، ٠، ٠، ٠،
١٢	٩، ٦، ٦، ٥، ٤، ٤، ٣، ٢، ١، ١، ١، ١، ١، ٠
١٣	٦، ٠، ٠
١٤	٦، ٢
١٥	٢

عينة من ١١٢ تلميذاً من تلاميذ المدارس الابتدائية ، حيث يلاحظ أن الوسط الحسابي $\overline{س}$ = ١٠٤,٥ وأن الانحراف المعياري ع = ١٦,٣ . وتفيـد قاعـدة تشيبيتشيف اذا طبقت على هذه البيانات أن ٧٥٪ على الأقل من المشاهدات تقع بـيـن $\overline{س}$ - ٢ ع ، $\overline{س}$ + ٢ع أي بـيــن ١٠٤,٥ - ٢(١٦,٣) ، ١٠٤,٥ + ٢(١٦,٣) أي بين (٧١,٩ ، ١٣٧,١) . على حين أنه يـلاحظ أن النسبـة الفعلية لعدد المشاهدات التي تقع داخل هذه الفترة تساوي $\frac{١٠٨}{١١٢}$ = ٠,٩٦

كذلك يلاحظ أن قاعدة تشبيتشيف تدل على وجود ٨٩٪ على الأقل من المشاهدات داخل الفترة سَ - ٣ع ، سَ + ٣ع أي بين ٥٦ ، ١٥٣ ، على حين أن النسبة الفعلية في البيانات تبلغ ١٠٠٪ .

٧ ـ وصف وتلخيص التوزيع التكراري ـ القاعدة العملية

يؤدي تطبيق قاعدة تشيبتشيف إلى نتائج محافظة في غالب الأحيان وذلك بسبب قابليتها للاستخدام في وصف جميع التوزيعات الاحصائية أياً كان شكلها . ويترتب على ذلك أنه إذا كان من الممكن إنشاء قواعد أخرى بديلة يقتصر تطبيقها على توزيعات تكرارية ذات شكل محدد فقط ، فإن نتائج تطبيق مثل هذه القواعد تكون أقل تحفظاً واكثر دقة . وتعتبر القاعدة العملية أكثر هذه القواعد أهمية وتستخدم اذا كان توزيع البيانات قريب من شكل المنحنى الطبيعي . وتوصف هذه القاعدة بالعملية لأنها تعكس الخبرة العملية للباحثين في المجالات المختلفة والتي أظهرت أن المنحنى الطبيعي يمثل تقريباً جيداً للعديد من مجموعات البيانات الإحصائية .

وتنص القاعدة العملية على ما يلي . إذا كان من الممكن افتراض أن شكل المدرج التكراري (المنحنى التكراري) لمجموعة المشاهدات يقترب من شكل المنحنى الطبيعي فإن :

(أ) حوالي ٦٨٪ من المشاهدات تقع داخل بعد قدره انحراف معياري واحد من الوسط الحسابي ، أي داخل الفترة (سَ - ع ، سَ + ع) .

(ب) حوالي ٩٥٪ من المشاهدات تقع داخل بعد قدره انحرافين معياريين من الوسط الحسابي ، أي داخل الفترة (سَ - ٢ع ، سَ + ٢ع) .

(جـ) حوالي ٩٩,٧٪ من المشاهدات تقع داخل بعد قدره ثلاث انحرافات معيارية من الوسط الحسابي ، أي داخل الفترة (سَ - ٣ع ، سَ + ٣ع) .

ويلاحظ أن القاعدة العملية تتميز عن قاعدة تشيبيتشيف بأنها تعطي قيماً

محددة للنسب وليس حداً أدنى لها فقط .

مثال(١٢): في دراسة عن تأثير سرعة السيارة على درجة الانسياب المروري على طريق معين ، سجلت سرعة ٥٠٠٠ سيارة مارة بهذا الطريق خلال أسبوع معين . وجد أن متوسط سرعة السيارة = ٨٠كم/ساعة وأن الانحراف المعياري للسرعة = ٢٠كم/ساعة . احسب النسب الآتية بافتراض أن المدرج التكراري للسرعات قريب في شكله من المنحنى الطبيعي :

(أ) نسبة السيارات التي تسير بسرعة تتراوح بين ٦٠ كم/ ساعة ، ١٠٠كم/ساعة .

(ب) نسبة السيارات التي تسير بسرعة تزيد عن ١٠٠كم في الساعة .

(حـ) نسبة السيارات التي تسير بسرعة تزيد عن ١٢٠كم في الساعة .

الحل :

(أ) يلاحظ أن سَ = ٨٠ ، ع = ٢٠ ، وأن الفترة (٦٠ ، ١٠٠) يمكن كتابتها على الشكل (٨٠ - ٢٠ ، ٨٠ + ٢٠) أي على الشكل (سَ - ع ، سَ +ع) . وتكون نسبة المشاهدات داخل هذه الفترة طبقاً للقاعدة العملية مساوية ٦٨٪ تقريباً .

(ب) لتحديد نسبة السيارات التي تسير بسرعة تزيد على ١٠٠كم/ساعة ، يلاحظ من الجزء (أ) أن ٦٨٪ من السيارات تسير بسرعة تتراوح بين ٦٠ ، ١٠٠كم . معنى ذلك أن ٣٢٪ من السيارات تسير بسرعة تقل عن ٦٠كم/ساعة أو تزيد عن ١٠٠كم/ساعة . ولما كان المنحنى الطبيعي متماثلاً فإن ذلك يعني تقسيم هذه النسبة بالتساوي على الجانبين . وبالتالي فإن نسبة السيارات التي تزيد سرعتها عن ١٠٠كم/ساعة تساوي $\frac{٣٢}{٢}$ = ١٦٪ تقريباً .

(حـ) باستخدام نفس أسلوب الإجابة عن (ب) ، يلاحظ أن ٩٥٪ من السيارات تتراوح سرعتها بين ٤٠ ، ١٢٠كم/ ساعة. معنى ذلك أن ٥٪ من السيارات تقل سرعتها عن ٤٠كم/ساعة أو تزيد سرعتها عن

١٢٠كم/ساعة . ويترتب على ذلك أن نسبة السيارات التي تزيد سرعتها عن ١٢٠كم/ساعة تساوي ٢,٥٪ ، بناءاً على تماثل المنحنى الطبيعي .

وتجدر الإشارة إلى أن اقتراب شكل توزيع المشاهدات من المنحنى الطبيعي يعتبر شرطاً أساسياً لصحة تطبيق القاعدة العملية . فمثلاً يلاحظ أن ١٠٠٪ من مجموعة المشاهدات { ١ ، ٢ } تقع داخل الفترة (س̄ - ع ، س̄ + ع) ، كذلك فإن ٣٣٪ من مجموعة المشاهدات { ١ ، ١ ، ٢ ، ٢ ، ٣ ، ٣ } تقع داخل الفترة (س̄ - ع ، س̄ + ع) ، وهي نتائج لا تتفق مع القاعدة العملية . وسوف نناقش بالتفصيل في الباب الثامن استخدامات المنحنى الطبيعي كتقريب للتوزيعات الاحصائية المختلفة ، وسوف يستفاد من ذلك في توضيح الشروط اللازمة لتطبيق القاعدة العملية وكيفية حساب نسب أخرى لوصف التوزيع التكراري . ما يهم الآن هو التأكيد على أنه في حالة اقتراب شكل التوزيع من المنحنى الطبيعي ، تكون الغالبية العظمى من المشاهدات (٩٥٪) واقعة داخل المدى (س̄ - ٢ع ، س̄ + ٢ع) ، كما تقع كل المشاهدات تقريباً داخل المدى (س̄ - ٣ع ، س̄ + ٣ع) . وعلى ذلك جرت العادة على اعتبار أي مشاهدة تبعد عن الوسط الحسابي بأكثر من ثلاث انحرافات معيارية مشاهدة شاذة أو متطرفة تتطلب دراسة خاصة لتحديد ما إذا كان هناك خطأ في تسجيلها ، أو ظروف خاصة أدت إلى نشأتها .

٨ ـ معامل الاختلاف

يقاس الانحراف المعياري بنفس وحدات المتغير محل الدراسة . ويترتب على ذلك عدم امكانية الاعتماد على هذا المقياس للمقارنة بين درجة تشتت توزيعين مختلفين في المواقف التي تكون فيها بيانات التوزيعين مقاسة بوحدات مختلفة . مثال ذلك المقارنة بين درجة التشتت في إيجارات المساكن في الاحياء المختلفة ودرجة التشتت في عدد السكان في هذه الأحياء ، أو المقارنة بين مجموعتين من بيانات الزمن احداها مسجلة بالساعات والأخرى

مسجلة بالدقائق . كذلك لا يمكن الاعتماد على الانحراف المعياري لمقارنة درجات التشتت اذا كانت هناك اختلافات واسعة في المتوسطات ، فمثلاً لا يمكن مقارنة التشتت في اجور العمال بالتشتت في اجور المديرين في احدى الشركات دون أخذ الاختلافات في متوسطات هذه الأجور في الاعتبار .

ويجب في مثل هذه المواقف حساب مقياس نسبي للتشتت يكون خالياً من وحدات القياس ويأخذ قيمة الوسط الحسابي في الاعتبار . ويعتبر معامل الاختلاف أهم هذه المقاييس وأكثرها استخداماً في التطبيقات المختلفة .

يعبر معامل الاختلاف عن الانحراف المعياري كنسبة مئوية من الوسط الحسابي ، أي أن :

$$\text{معامل الاختلاف} = \frac{\text{الانحراف المعياري}}{\text{الوسط الحسابي}} \times ١٠٠$$

وتستخدم كلمة معامل للدلالة على أن المقياس الناتج لا يتأثر بوحدات القياس ولا يكون له تمييزاً . وفيما يلي بعض الأمثلة على استخدامات معامل الاختلاف .

مثال (١٣) : يراد مقارنة درجة التأرجح في مبيعات شركتين أ ، ب بين فترة زمنية وأخرى . جمعت بيانات عن قيمة المبيعات الأسبوعية للشركة أ فوجد أن متوسط قيمة هذه المبيعات خلال العام الماضي = ٥٢ ألف درهم وأن الانحراف المعياري = ١١ ألف درهم . وجمعت بيانات عن قيمة المبيعات الشهرية للشركة ب فوجد أن متوسط قيمة هذه المبيعات خلال العام الماضي = ٢٠٠ ألف درهم وأن الانحراف المعياري = ١٨ ألف درهم .

لا يصلح الانحراف المعياري في هذه الحالة كأساس لمقارنة درجة التشتت في مبيعات الشركتين ، وذلك لأن قيم الشركة أ هي مبيعات أسبوعية بينما تمثل قيم الشركة ب مبيعات شهرية.وقد انعكس ذلك في كبر متوسط قيم مبيعات الشركة ب وفي كبر الانحراف المعياري لهذه القيم . ويجب الاعتماد في المقارنة ، بدلاً من ذلك ، على معامل الاختلاف ، حيث يلاحظ أن :

معامل الاختلاف لمبيعات الشركة أ = $\frac{١١}{٥٢}$ × ١٠٠ = ٢١,٢٪

معامل الاختلاف لمبيعات الشركة ب = $\frac{١٨}{٢٠٠}$ × ١٠٠ = ٩,٠٪ .

وبناء على ذلك ، يمكن القول أن المبيعات الشهرية للشركة ب كانت أقل تغيراً بشكل نسبي من المبيعات الأسبوعية للشركة ب . لاحظ أن الاعتماد على الانحرافات المعيارية في هذه الحالة يؤدي إلى نتائج عكسية .

مثال (١٤) : ترغب احدى شركات الشحن في مقارنة درجة التشتت في احجام الصناديق المستخدمة للشحن وفي أوزان هذه الشحنات . وجد أن متوسط حجم الصندوق = ١٢,٦ قدم^{٣} وأن الانحراف المعياري = ٢,١ قدم مكعب . كذلك وجد أن متوسط وزن محتويات الصندوق = ٢٨,٤ كجم وأن الانحراف المعياري = ٨,١ كجم .

لا يمكن استخدام الانحرافات المعيارية لاجراء المقارنة في هذه الحالة لأن وحدات القياس مختلفة . ويمكن الاعتماد ، بدلاً من ذلك ، على معامل الاختلاف ، حيث يلاحظ أن :

معامل اختلاف الحجم = $\frac{٢,١}{١٢,٦}$ × ١٠٠ = ١٦,٧٪

معامل اختلاف الوزن = $\frac{٨,١}{٢٨,٤}$ × ١٠٠ = ٢٨,٥٪ .

ويعني ذلك أن الاختلافات النسبية في أحجام الصناديق أصغر من الاختلافات النسبية في أوزان محتويات هذه الصناديق .

ويجب الإشارة إلى أن معامل الاختلاف يستخدم عادة عندما تكون جميع المشاهدات موجبة مثل المشاهدات عن المبيعات أو الحجم أو الوزن أو الطول . إذا كانت هناك قيماً سالبة وموجبة في البيانات فإن استخدام معامل الاختلاف قد يؤدي الى حدوث لبس وينصح بعدم الاعتماد عليه في هذه الحالة .

٩ ـ الوحدات المعيارية

يرغب الطالب عادة في مقارنة درجته التي يحصل عليها في امتحان ما بالدرجات التي حصل عليها الطلبة الآخرون . ويهدف الطالب من هذه المقارنة بتحديد مكانه النسبي بين زملائه ويشمل ذلك معرفة ما إذا كانت درجته تقع فوق المتوسط أم لا وما إذا كانت درجته تدخل ضمن أفضل ٥٪ أو ١٠٪ أو١٥ ٪ . . . من الدرجات . ويتم ذلك بالاعتماد على المقاييس التي تصف الوضع النسبي لكل مشاهدة داخل مجموعة المشاهدات .

حصل أحد الطلبة على درجة ٧٢ في مساق اللغة الانجليزية وعلى درجة ٦٨ في مساق الرياضيات . يراد معرفة ما إذا كان وضع الطالب النسبي بين طلبة مساق اللغة الانجليزية أفضل أم أسوأ من وضعه النسبي بين طلبة مساق الرياضيات . كانت قيمة الوسط الحسابي للدرجات في اللغة الانجليزية = ٦٠ والانحراف المعياري = ١٦ وكانت قيمة الوسط الحسابي للدرجات في الرياضيات يساوي ٥٨ والانحراف المعياري = ١٠ .

يلاحظ أن درجة الطالب في اللغة الانجليزية تزيد عن المتوسط بمقدار ٧٢ ـ ٦٠ = ١٢ درجة . ولما كان الانحراف المعياري المناظر = ١٦ فإن ذلك يعني أن درجة الطالب تزيد عن المتوسط بمقدار $\frac{١٢}{١٦}$ = ٧٥, انحرافاً معيارياً . وبنفس الطريقة ، تزيد درجة الطالب في الرياضيات عن المتوسط بمقدار $\frac{٦٨ - ٥٨}{١٠}$ = ١ إنحرافاً معيارياً . ويعني ذلك أنه على الرغم من أن درجة الطالب في اللغة الانجليزية أعلى من درجته في الرياضيات إلا أن وضعه النسبي في مساق الرياضيات أفضل من وضعه النسبي بين الطلبة في مساق اللغة الانجليزية .

إذا تم قياس انحرافات المشاهدات المختلفة عن الوسط الحسابي كنسبة من الانحراف المعياري فإن القيم الناتجة تسمى وحدات معيارية . وبشكل عام ، إذا كانت س ترمز لقيمة مشاهدة ما من مجموعة مشاهدات وسطها

الحسابي هو $\bar{س}$ وانحرافها المعياري هو ع فإن الوحدة المعيارية المناظرة ويرمز لها بالرمز ى هي :

$$ى = \frac{س - \bar{س}}{ع}$$

وتقيس قيمة ى المسافة بين المشاهدة والوسط الحسابي بوحدات من الانحراف المعياري . وقد تكون قيمة ى موجبة أو سالبة تبعاً لما إذا كانت قيمة المشاهدة س أكبر من المتوسط أو أقل منه .

مثال (١٥) : يريد شخصان تخرجا حديثاً من الجامعة أحدهما طبيب والآخر مدرس مقارنة عرض العمل الذي تلقاه كل منهما . عرض على الطبيب عمل بمبلغ ٨٠٠٠ درهم شهرياً وعرض على المدرس عمل بمبلغ ٦٠٠٠ درهم شهرياً . كان متوسط أجر الطبيب الجديد في هذا العام ٩٠٠٠ درهم شهرياً والانحراف المعياري ٩٠٠ درهم ، بينما كان متوسط أجر المدرس الجديد في هذا العام ٥٥٠٠ درهم شهرياً والانحراف المعياري ٥٠٠ درهم . والمطلوب تحديد أي العرضين أفضل نسبياً .

الحل :

الدرجة المعيارية لأجر الطبيب : $ى = \frac{٨٠٠٠ - ٩٠٠٠}{٩٠٠} = - \frac{١٠٠٠}{٩٠٠} = -١,١$

(أي أن ٨٠٠٠ تقل عن المتوسط بمقدار ١,١ انحرافاً معيارياً) .

الدرجة المعيارية لأجر المدرس $ى = \frac{٦٠٠٠ - ٥٥٠٠}{٥٠٠} = ١$

(أي أن ٦٠٠٠ تزيد عن المتوسط بمقدار انحراف معياري واحد) .

ويستنتج من ذلك أن العرض الذي تلقاه المدرس أفضل نسبياً (أي بالنسبة لزملائه المدرسون) من العرض الذي تلقاه الطبيب (بالنسبة لزملائه الأطباء) .

وتجدر الإشارة إلى أن الوحدات المعيارية تكون مفيدة اذا كان شكل التوزيع يقترب من المنحنى الطبيعي . اذ يمكن القول بناءاً على القاعدة العملية في هذه الحالة أن نسبة قيم ى التي تقع خارج المدى (- ٢ ، ٢) تساوي ٥٪ فقط ، كذلك تقع جميع قيم ى تقريباً داخل المدى (ـ ٣ ، ٣) . اذا وجدت إحدى قيم ى خارج هذا المدى ، فإن ذلك يعني غالباً أن المشاهدة المناظرة شاذة أو متطرفة .

١٠ ـ معامل الالتواء

يقصد بالالتواء عدم التماثل في شكل التوزيع التكراري . وقد سبقت الإشارة إلى أن درجة واتجاه الالتواء يعتبر سمة أساسية من سمات التوزيع التكراري . وتعتمد المقاييس البسيطة للالتواء على العلاقة بين قيمة المنوال وقيمة الوسيط وقيمة الوسط الحسابي للتوزيع . ذلك أنه من المعروف أن هذه القيم تكون متساوية اذا كان التوزيع التكراري متماثلاً . وتكون قيمة المنوال أقل من قيمة الوسيط وكلاهما أقل من قيمة الوسط الحسابي اذا كان التوزيع التكراري ملتوياً الى اليمين . وعلى العكس تكون قيمة المنوال أكبر من قيمة الوسيط وكلاهما أكبر من قيمة الوسط الحسابي اذا كان التوزيع التكراري ملتوياً الى اليسار .

وعلى ذلك يمكن تعريف معامل للالتواء كما يلي :

$$\text{معامل الالتواء} = \frac{\text{٣ (الوسط الحسابي ـ الوسيط)}}{\text{الانحراف المعياري}} \ .$$

وقد استخدم الوسيط بدلاً من المنوال في هذا المعامل لصعوبة تحديد قيمة المنوال بدقة في التوزيعات التكرارية . أما القسمة على الانحراف المعياري فتهدف الى التخلص من وحدات القياس حتى يمكن استخدامه للمقارنة بين التوزيعات المختلفة . وتساوي قيمة هذا المعامل صفراً إذا كان التوزيع متماثلاً (الوسط الحسابي = الوسيط) وتكون قيمته موجبة اذا كان

الالتواء في الاتجاه الموجب (الوسط الحسابي أكبر من الوسيط) وتكون قيمته سالبة إذا كان الالتواء في الاتجاه السالب (الوسط الحسابي أقل من الوسيط) .

وتجدر الإشارة إلى أن هذا المعامل لا يستخدم كثيراً في التطبيقات العملية . ويكفي عادة في المراحل الأولية لتحليل البيانات بوصف الالتواء من خلال رسم شكل التوزيع أو بدراسة العلاقة بين مقاييس المركز المختلفة ، وذلك بدلاً من الاعتماد على مقياس وحيد قد يكون من الصعب تفسير معنى قيمته .

هناك مقاييس أخرى للالتواء تكون أكثر دقة وأسهل تفسيراً . وتتطلب هذه المقاييس معرفة أساليب إحصائية رياضية لا يتسع المقام هنا لذكرها .

تمرينات

(أ) اذا كان هناك عينة مشاهداتها هي (١ ، ٢ ، ١ ، ٢ ، ٤) وعينة أخرى مشاهداتها هي (١ ، ٢ ، ١ ، ٢ ، ٩) فاذكر بدون اجراء حسابات ما إذا كان الانحراف المعياري في العينة الأولى سيكون أكبر من الانحراف المعياري في العينة الثانية أم لا مع توضيح سبب اجابتك . تأكد من صحة الاجابة بحساب الانحراف المعياري في كل حالة .

(ب) اذا كان هناك عينة من ثلاث مشاهدات وكان انحراف مشاهدتين منهما عن الوسط الحسابي يساوي − ٥ ، − ٢ على الترتيب فما هي قيمة انحراف المشاهدة الثالثة عن الوسط الحسابي ؟ . احسب قيمة كل من الانحراف المتوسط والانحراف المعياري لبيانات هذه العينة .

(حـ) احسب قيمة الانحراف المعياري لبيانات عينة من ١٠ مفردات اذا علم أن مجـ س = ٢٠ ، مجـ $س^2$ = ١٦١ .

(ى) احسب قيمة التباين لعدد الأطفال في الأسرة في كل من العينات الثلاث التالية وقارن بين درجة التشتت في هذه العينات :

العينة الأولى : ٣ ، ٢ ، ١ ، ٥ ، ٦

العينة الثانية : ٣ ، ٢ ، ١ ، ٤ ، ٦

العينة الثالثة : ٣ ، ٢ ، ١ ، ٤ ، ٥

(هـ) احسب قيمة كل من المدى المطلق والمدى الربيعي للبيانات الآتية التي تمثل أوزان ١١ تلميذاً بالكيلوجرام :

٦٢ ، ٢٣ ، ٢٧ ، ٥٦ ، ٥٢ ، ٣٤ ، ٤٢ ، ٤٠ ، ٦٨ ، ٥٤ ، ٨٣ .

(٢) (أ) اعط مثالاً لمجموعتي بيانات تتألف كل منهـا من ٥ مشاهـدات يكون لهما نفس الـوسط الحسابي ويختلفـان في قيمة الانحـراف المعياري .

(ب) اعط مثالاً لمجموعتي بيانات تتألف كل منها من ٥ مشاهدات يكون لهمـا نفس قيمة الانحـراف المعياري ويختلفـان في قيمة الـوسط الحسابي .

(حـ) احسب قيمـة كـل من الـوسط الحسـابي والانحــراف المعيـاري للبيانات الآتية التي تمثل عينـة من ١١ فرداً : ٥٢ ، ١٣ ، ١٧ ، ٤٦ ، ٤٢ ، ٢٤ ، ٣٢ ، ٣٠ ، ٥٨ ، ٤٤ ، ٧٣ .

(د) اطرح ١٠ من كل قيمة من القيم المعطاة في (جـ) واحسب الوسط الحسابي والانحراف المعـراي للقيم الناتجـة . قارن بين النتـائج التي تحصل عليها والنتائج التي حصلت عليها في (جـ) .

(هـ) اضـرب كل قيمـة من القيم المعـطاة في (جـ) في ١٠ واحسب الوسط الحسابي والانحراف المعياري للقيم النـاتجة . قـارن بين النتائج التي تحصل عليها والنتائج التي حصلت عليها في (جـ) .

(٣) فيما يلي أطوال عينة من ٢٨ تلميذاً لأقرب سنتمتر :

٧٦	٧٧	٤٣	٥٩	٥٠	١٠٥	٧٧	٩٥	١٢٠	١٢٦
٥٥	٨٣	٨٢	١٣٢	١٢٦	١٣٦	١٤١	١٣٥	١١٥	١٢٠
١٠٤	١٠٨	١٣١	١٢٣	١٠٤	١٣٠	١٤١	١٥١		

(أ) احسب قيمة الوسط الحسابي لهذه البيانات .

(ب) احسب انحراف كل مشـاهدة عن الـوسط الحسابي وتـأكد من أن مجموع هذه الانحرافات يساوي الصفر .

(حـ) احسب قيمة الانحراف المتوسط ، وفسر معناه .

(د) احسب قيمة كل من المدى المطلق والمدى الربيعي لهذه البيانات .

(هـ) ارسم شكل الصندوق والشعيـرات لهذه البيـانات وعلق على نمط الاختلاف الذي يظهره هـذا الشكل .

(٤) يعطي الجدول التالي التـوزيع العمـري لعينة من ١٠٠٠ شخص ممن اشتركوا في الـرحلات السياحية التي نظمتها إحدى الشركات خلال هذا العام .

فئات العمر بالسنوات	٢٠ ـ ٢٤	٢٥ ـ ٢٩	٣٠ ـ ٣٤	٣٥ ـ ٣٩
عدد الأشخاص	١٢٩	٢٢١	٣١٠	١٦٣

فئات العمر بالسنوات	٤٠ ـ ٤٤	٤٥ ـ ٤٩	٥٠ ـ ٥٤	المجموع
عدد الأشخاص	١٠٥	٦٢	١٠	١٠٠٠

(أ) احسب قيمة الانحراف المتوسط للعمر وفسر معناه .

(ب) احسب قيمة كل من التباين والانحراف المعياري للعمر .

(جـ) احسب قيمة كل من الوسيط والربيع الأول والربيع الثالث للعمر .

(د) ارسم شكل الصندوق والشعيرات المناظر لهذه البيانات ، وعلق على نمط الاختلاف الذي يظهره هذا الشكل .

(هـ) ارسم المدرج التكراري المناظر. هـل يقدم المـدرج وصفاً لنمط الاختلاف أفضل من ذلك الذي يظهر في الجزء (د) أم لا ؟ وضح سبب إجابتك .

(٥) طلب من كل تلميذ في عينة من ١٥٠ تلميذاً القيام بترجمة قطعة أدبية من اللغة الانجليزية إلى اللغة العربية . فيمـا يلي التوزيـع التكراري لعـدد الأخطاء التي ارتكبها كل تلميذ .

عدد الأخطاء	١٧ ـ ١٩	٢٠ ـ ٢٢	٢٣ ـ ٢٥	٢٦ ـ ٢٨	٢٩ ـ ٣١	٣٢ ـ ٣٤	المجموع
عدد التلاميذ	٥	٦٣	٣٩	٢٤	١٧	٢	١٥٠

(أ) احسب قيمة كل من الوسط الحسابي والانحراف المعياري .

(ب) احسب قيمة كل من الوسيط والمدى الربيعي .

(جـ) ارسم شكل الصندوق والشعيرات المناظر لهذه البيانات ، وعلق على نمط الاختلاف كما يظهر في هذا الشكل .

(د) ارسم المدرج التكراري المناظر .

(٦) فيما يلي التوزيع التكراري لعدد الحوادث الأسبوعية التي وقعت عند احد التقاطعات المرورية خلال الأسابيع الخمسين الماضية.

عدد الحوادث	صفر	١	٢	٣	٤	٥	المجموع
عدد الأسابيع	٢٠	١٢	٨	٥	٣	٢	٥٠

أ ـ بين أن مجموع انحرافات المشاهدات عن وسطها الحسابي يساوي الصفر .

ب ـ احسب قيمة الانحراف المتوسط لعدد الحوادث الأسبوعية وفسر معنى النتيجة التي تحصل عليها .

جـ ـ احسب قيمة الانحراف المعياري لعدد الحوادث الاسبوعية .

(٧) (أ) اذا كانت هناك المشاهدات ١٤ ، ١٧ ، ١٣ ، ٤١ ، ١٢ ، فإحسب قيمة المجموع مجـ | س ـ الوسيط| . احسب قيمة مجـ |س ـ أ| لمجموعة تختارها من قيم أ وبين أن المجاميع في هذه الحالة تكون دائماً أكبر من مجـ | س ـ الوسيط| .

(ب) يمكن تفسير معنى الانحراف المعياري بدلالة المسافات الثنائية بين المشاهدات المختلفة . فمثلاً إذا كانت هناك المشاهدات (٥ ، ٦ ، ٧ ، ٨ ، ٩) فإن هناك فروقاً ثنائية بين هذه المشاهدات عددها ١٠ هي (٦ ـ ٥) ، (٧ ـ ٥) ، (٨ ـ ٥) ، . . . ، (٩ ـ ٥) . إذا ربعت هذه الفروق وأخذ متوسط هذه المربعات فإن الناتج يكون ضعف قيمة التباين . تأكد من صحة هذه القاعدة باستخدام البيانات المعطاة. اشرح دلالة هذه النتيجة عند تفسير

معنى الانحراف المعياري كمقياس للاختلافات بين المشاهدات .

(٨) فيما يلي طول وعرض الجمجمة في عينة من ٢٥ ذكراً من سكـان إحدى الدول ، حيث يلاحظ بالطبع أن متوسط الطول أكبر من متوسط العرض . هل تباين الطول أكبر من تباين العرض أيضاً ؟

الطول بالميلمتر : ١٩١ ١٩٥ ١٨١ ١٨٣ ١٧٦ ٢٠٨ ١٨٩ ١٩٧ ١٨٨ ١٩٢ ١٧٩
العرض بالميلمتر : ١٥٥ ١٤٩ ١٤٨ ١٥٣ ١٤٤ ١٥٧ ١٥٠ ١٥٩ ١٥٢ ١٥٠ ١٥٨
الطول بالميلمتر : ١٨٣ ١٧٤ ١٩٠ ١٨٨ ١٦٣ ١٩٥ ١٨١ ١٨٦ ١٧٥ ١٩٢ ١٧٤
العرض بالميلمتر : ١٤٧ ١٥٠ ١٥٩ ١٥١ ١٣٧ ١٥٥ ١٥٣ ١٤٥ ١٤٠ ١٥٤ ١٤٣
الطول بالميلمتر : ١٧٦ ١٩٧ ١٩٠
العرض بالميلمتر : ١٣٩ ١٦٧ ١٦٣

(٩) ارسم شكل الصندوق والشعيرات لكل من بيانات الطول وبيانات العرض في التمرين السابق وقارن بين نمط الاختلاف في كل منهما .

(١٠) قارن بين درجة التشتت في مستوى الدخل السنوي في البلدان أ ، ب ، حـ باستخدام البيانات الآتية :

البلد	متوسط الدخل السنوي	الانحراف المعياري للدخل السنوي
أ	١٠٠٠٠ دولار	٥٠٠٠ دولار
ب	٣٠٠٠ جنيه	١٥٠٠ جنيه
جـ	٣٦٠٠٠ درهم	٣٦٠٠٠ درهم

(١١) (أ) إعط مثالاً لمجموعة بيانات يفضل استخدام المدى الربيعي لقياس درجة التشتت فيها بدلاً من الانحراف المعياري . اشرح سبب هذا التفضيل .

(ب) اذا كـانت هناك المشاهدات ٥ ، ٦ ، ٧ ، ٨ ، ٩ ، احسب

المجمـوع مجـ (س ـ الوسط الحسابي)[٢] . احسب كذلك المجاميع مجـ (س ـ أ)[٢] لأي قيم تختارها للثابت أ وبين أن هذه المجـاميع تكـون دائماً أقـل من مجـ (س ـ الوسط الحسابي)[٢] .

(١٢) فيما يلي التوزيع التكراري لمبلغ الحوافز المدفوع لكل بائع في عينة من ٦٥ بائعاً في إحدى الشركات وذلك خلال عام ١٩٨٥ .

قيمة الحوافز بالدرهم	٠ ـ ٩٩٩	١٠٠٠ ـ ١٩٩٩	٢٠٠٠ ـ ٣٩٩٩
عدد البائعين	٤	١٧	٢٧

قيمة الحوافز بالدرهم	٤٠٠٠ ـ ٥٩٩٩	٦٠٠٠ ـ ٩٩٩٩	١٠٠٠٠ ـ ١٩٩٩٩	المجموع
عدد البائعين	١١	٥	١	٦٥

أ ـ احسب قيمة الوسط الحسابي للحوافـز المدفـوعة للبـائع خـلال عام ١٩٨٥ . ما هي القيمة الكلية للحوافز المدفوعة للبـائعين في هذه العينة .

ب ـ اذا كانت الشركة بها ٧٠٠ عامل ، احسب تقـديراً لقيمـة الحوافـز الكلية المدفوعة للبائعين في هذه الشركة .

حـ ـ احسب كلاً من الانحراف المعيـاري والتباين للحـوافز المـدفوعـة للبائع خلال عام ١٩٨٥ . ماهي وحدات القياس لكل منهما .

د ـ اذا علم أن قيمة الانحراف المعياري للحوافز المدفوعة لنفس مفردات هذه العينة في عام ١٩٨٣ بلغ ١٥١٧ درهم . ما هو التغير الـذي حدث في شكل توزيع الحوافز خلال العامين ١٩٨٣ ـ ١٩٨٥ .

١٣ ـ في دراسة عن ٣٨٥٠ عامـلاً باليـومية في إحـدى المدن وجـد أن الحد الأقصى لـلأجـر الشهـري لـلعـامـل = ٢٨٣٠,٥٠ درهم والـحـد الأدنى = ٢٦٣,٤٠ درهم وأن الوسط الحسابي = ١٧٣٠,٥٠ درهم وأن

الـوسيط =١٥٦٤,٤٠ درهم وأن الانحـراف المـعيـاري =٢١٥,٢٥ درهم .

(أ) استخدم قاعدة تشيبيتشيف لتقدير نسبة العمـال الذين يحصلون على أجور شهرية تقل عن ١٣٠٠ درهم أو تزيد عن ٢١٦١ درهم ؟

(ب) قـدر نسبـة العمـال الـذين تقـع أجـورهم بين ٨٦٩,٥٠ درهم ، ٢٥٩١,٥٠ درهم .

(حـ) ما هو اتجاه الالتواء في هذه البيانات . اشرح سبب إجابتك .

١٤ ـ وجد في نفس الدراسة السابقة في التمرين (١٣) أن متوسط عدد الأيام التي يشتغلها العامل =١٧,٣ يوم وأن الـوسيط =١٦,٢ يوم والانحـراف المعياري = يوم واحد . يراد المقارنة بين درجة التشتت في توزيع الأجور ودرجة التشتت في توزيع الأيام المشتغلة . استخدم مقياساً مناسباً لاجراء هذه المقارنة .

١٥ ـ فيما يلي بعض خصائص توزيع الدخل في بلدين أ ، ب . احسب قيمة معامل الالتواء في توزيع الدخل لكل من البلدين وعلق على النتائج .

البلد	الوسط الحسابي	الوسيط	الانحراف المعياري
أ	٤٩٣٥٠ درهم	٤٥٦٠٠ درهم	٨٧٠٠ درهم
ب	٥٠٩٤٠ درهم	٤٦٥٠٠ درهم	٧٥٠٠ درهم

١٦ ـ في دراسة عن الزمن الذي يقضيه الطالب في مطعم الجامعة أثناء تناوله طعام العشاء ، وجد أن الوسط الحسابي لهذا الزمن = ٣٥ دقيقة وأن الانحراف المعياري = ٥ دقائق .

(أ) بدون افتراض شكل معين للتوزيع التكراري ، قدر نسبة الطلبة الذين يقضون زمناً يتراوح بين ٢٥ دقيقة ، ٤٥ دقيقـة أثناء تنـاولهم طعام العشاء .

(ب) بدون افتراض شكل معين للتوزيع التكراري ، قدر نسبة الطلبة الذين يقضون زمناً يقل عن ٢٠ دقيقة أو يزيد عن ٥٠ دقيقة .

(جـ) اذا افترض أن شكل التوزيع يقترب من المنحنى الطبيعي فاحسب نسبة الطلبة الذين يقضون زمناً يتراوح بين ٢٥ ، ٤٥ دقيقة ، ونسبة الطلبة الذين يقضون زمناً يقل عن ٢٠ دقيقة .

١٧ ـ إذا كان معلوماً أن الوسط الحسابي للدرجات التي حصل عليها الطلبة في أحد الامتحانات يساوي ٤٥ وأن الانحراف المعياري = ٧ . احسب الدرجات المعيارية المناظرة للدرجات التالية : ٣٢ ، ٤٧,٥ ، ٤٢ ، ٦١ . فسر معنى النتائج التي تحصل عليها .

١٨ ـ ينتمي شخصان أ ، ب إلى مجموعتين عمريتين مختلفتين . اذا كان الوسط الحسابي لوزن مفردات المجموعة الأولى = ٧٣ كجم والانحراف المعياري =٧ كجم وكان الوسط الحسابي لوزن مفردات المجموعة الثانية = ٨٠ كجم والانحراف المعياري =٨,٥ كجم ، وإذا علم أن وزن الشخص أ يساوي ٨٩ كجم ووزن الشخص ب يساوي ٩٦,٥ كجم فحدد أي الشخصين يعتبر أكثر سمانة بالنسبة لمجموعته العمرية .

١٩ ـقيس ضغط دم الشخص أ يومياً لعدة أسابيع فوجد أن المتوسط =٢٠٢ وأن الانحراف المعياري =١٢,٥ . وقيس ضغط الدم للشخص ب بنفس الأسلوب فوجد أن المتوسط = ١٢٤ والانحراف المعياري =٨,١ . قارن بين درجة التشتت في قراءات ضغط الدم للريضين .

٢٠ ـ يتألف الطلبة الذين يدرسون الاحصاء هذا العام والبالغ عددهم ٥٠٠ من ٣٠٠ طالب ، ٢٠٠ طالبة . وجد أن متوسط طول الطالب = ١٧٥ سم وأن الانحراف المعياري لطول الطالب = ٨ سم . كذلك وجد أن متوسط طول الطالبة =١٦٥سم وأن الانحراف المعياري لطول الطالبة = ٨ سم .

(أ) احسب الوسط الحسابي لطول الطلبة عموماً في هذه المجموعة .

(ب) احسب قيمة الانحراف المعياري لطول الطلبة عموماً .

(جـ) استخدم قاعدة تشيبيتشيف لتقدير نسبة الطلبة الذين تتراوح أطوالهم بين ١٦٥ ، ١٧٧ سم .

٢١ ـ يعطي الجدول التالي التوزيع التكراري لعدد مرات الغياب أثناء الفصل الدراسي لعينة من ١٠٠ طالب .

عدد مرات الغياب	صفر	١	٢	٣	٤	٥	٦	٧	٨	المجموع
عدد الطلبة	٤	٨	٢١	٤٥	١٢	٤	صفر	١	٥	١٠٠

أ ـ احسب قيمة الوسط الحسابي والانحراف المعياري لهذه البيانات .

ب ـ استخدم القاعدة العملية لحساب نسبة الطلبة الذين تتراوح عدد مرات غيابهم بين (س − ع ، س + ع) ، وبين (س − ٢ ع ، س + ٢ ع) .

حـ ـ احسب نسبة الطلبة الذين تتراوح عدد مرات غيابهم بين (س − ع ، س + ع) ، وبين (س − ٢ ع ، س + ٢ ع) مباشرة من الجدول التكراري . قارن بين هذه النتائج وبين النتائج في (ب) ومن ثم علق على مدى جودة القاعدة العملية في تمثيل هذا التوزيع .



الباب الثامن

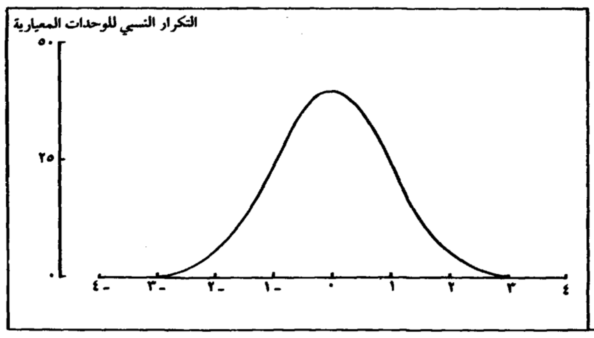
استخدام المنحنى الطبيعي لوصف البيانات

ناقشنا فيما سبق كيفية الاعتماد على المدرج التكراري لوصف السمات العامة لنمط الاختلاف في المشاهدات . وناقشنا كذلك امكانية استخدام عدد من المقاييس العددية ، وبصفة خاصة الوسط الحسابي والانحراف المعياري كبديل لتحقيق نفس الغرض . وأكدنا على أن نجاح هذه المقاييس العددية في وصف نمط الاختلاف يعتمد على الشكل العام للبيانات المستخدمة . فمثلاً ، يعتبر كلاً من الوسط الحسابي والانحراف المعياري مقاييس وصفية جيدة عندما يقترب الشكل العام للتوزيع التكراري من شكل المنحنى الطبيعي . وعلى ذلك ، ينصح دائماً بدراسة الشكل العام للتوزيع التكراري للمشاهدات كخطوة أولى عند التحليل الاحصائي . ويتم في ضوء هذه الدراسة استخدام وتفسير المقاييس العددية المختلفة بشكل ملائم .

سبقت الإشارة كذلك إلى أن أساس القاعدة العملية لوصف التوزيعات التكرارية باستخدام الوسط الحسابي والانحراف المعياري يكمن فيما لاحظه الباحثون من أن التوزيع التكراري للكثير من الظواهر التي تنشأ في الحياة العملية يقترب من شكل المنحنى الطبيعي . ويعتبر المنحنى الطبيعي تبعاً لذلك مفهوماً إحصائياً أساسياً . وقد اقترح الاحصائي كوتيليت منذ عام ١٨٧٠ الاعتماد على هذا المنحنى كمعيار تقارن به المدرجات التكرارية لمجموعات البيانات المختلفة .

١ ـ المنحنى الطبيعي

يعطي شكل (١) مثالاً لمنحنى طبيعي . ويرسم هذا المنحنى باستخدام معادلة رياضية معينة . ولن تذكر هذه المعادلة هنا لدواعي البساطة ، ذلك أنه يمكن دراسة واستخدام الخصائص الرياضية لهذا المنحنى دون اللجوء مباشرة إلى المعادلة . ويتسم هذا المنحنى بأنه متماثل حول الصفر ، ويقع كله فوق المحور الأفقي ، وأن المساحة المحصورة تحته خارج المدى (ـ ٤ ، ٤) صغيرة جداً بحيث يبدو أن المنحنى يبدأ في ملامسة المحور الأفقي بين ٣ ، ٤ . وسوف نهتم بكيفية حساب المساحات المختلفة تحت هذا المنحنى وذلك باستخدام جداول احصائية محسوبة فعلاً .



شكل (١) : المنحنى الطبيعي

ويقترب الشكل العام لتوزيع كثير من المتغيرات الاحصائية من شكل هذا المنحنى الطبيعي ، وذلك بشرط تحويل المشاهدات الى وحدات معيارية قبل رسمها . ويتطلب ذلك كما هو معروف طرح قيمة الوسط الحسابي من كل مشاهدة ، ثم قسمة الناتج على قيمة الانحراف المعياري . ويعني ذلك أن عملية المطابقة بين التوزيع التكراري المشاهد والمنحنى الطبيعي تعتمد فقط

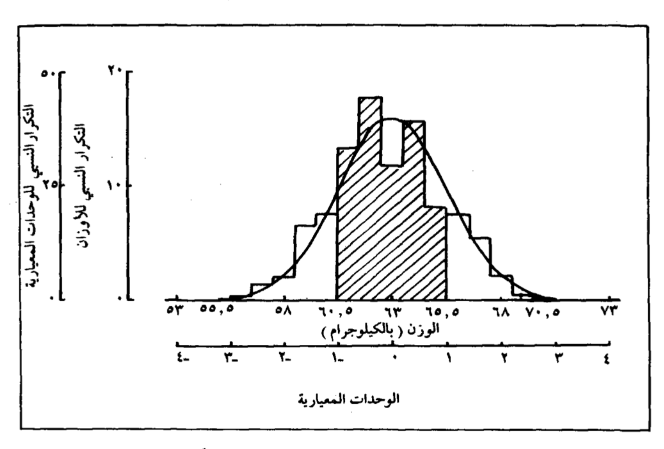
على الوسط الحسابي والانحراف المعياري للمشاهدات . ويمكن القول بناءاً على ذلك أن الوسط الحسابي والانحراف المعياري يعطيان معلومات كافية لوصف التوزيع التكراري وصفاً كاملاً إذا كان شكل هذا التوزيع قريب من المنحنى الطبيعي . وتتضح هذه المفاهيم الأساسية من الأمثلة التالية .

مثال (١) : جمعت بيانات عن الوزن في عينة من النساء فوجد أن الوسط الحسابي لوزن المرأة يساوي ٦٣كجم وأن الانحراف المعياري = ٢,٥كجم . يعطي شكل (٢) المدرج التكراري النسبي المناظر لهذه البيانات مقارناً بالمنحنى الطبيعي . ويلاحظ وجود زوجين من المحاور في الشكل . يمثل الزوج الداخلي مقياس الرسم اذا استخدمت الأوزان الأصلية لرسم الشكل ، بينما يمثل الزوج الخارجي مقياس الرسم اذا استخدمت الوحدات المعيارية المناظرة للرسم فمثلاً يلاحظ على المحاور الأفقية أن ٦٣ كجم تناظر وحدة معيارية قدرها صفر وأن $\frac{1}{2}$ ٦٥ كجم تناظر وحدة معيارية قدرها ١ وأن ٥٨ تناظر وحدة معيارية قدرها ـ ٢ وهكذا .

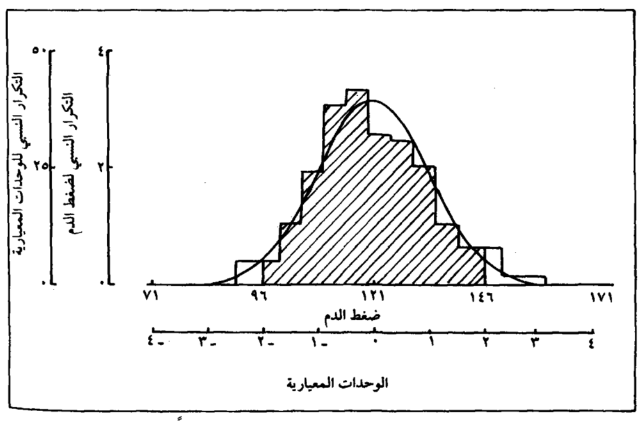
ويمكن حساب المساحات تحت المدرج التكراري أو تحت المنحنى الطبيعي بالاعتماد على أي زوج من هذه المحاور . وفي جميع الحالات تكون المساحة الكلية تحت المدرج التكراري النسبي مساوية للمساحة الكلية تحت المنحنى الطبيعي وكلاهما يساوي الواحد الصحيح .

أشرنا عند الحديث عن القاعدة العملية في الباب السابق أن حوالي ٦٨٪ من المشاهدات تقع داخل بعد قدره انحراف معياري واحد من الوسط الحسابي . ولتوضيح سبب ذلك يلاحظ أن نسبة النساء في شكل (٢) الذين تتراوح أوزانهم بين س ـ ع ، س + ع تمثل المساحة تحت المدرج التكراري المحصورة بين ـ ١ ، + ١ وحدة معيارية ، وهي المساحة المظللة في الشكل .

يتضح من الشكل أيضاً وجود تقارب بين شكلي المدرج التكراري والمنحنى الطبيعي وذلك على الرغم من وجود بعض النتوءات . ويلاحظ وجود



شكل (٢) : المدرج التكراري النسبي لأوزان النساء مقارناً بالمنحنى الطبيعي



شكل (٣): المدرج التكراري النسبي لضغط دم السكان مقارناً بالمنحنى الطبيعي

أجزاء صغيرة من المدرج خارج المنحنى ووجود أجزاء صغيرة لا تنتمي للمدرج داخل المنحنى وأن هذه الأجزاء توازن بعضها بعضاً بحيث أن المساحة المظللة تساوي تقريباً المساحة تحت المنحنى الطبيعي بين − ١ ، + ١ . وسنرى فيما بعد أن خصائص المنحنى الطبيعي تشير إلى أن هذه المساحة = ٠,٦٨ .

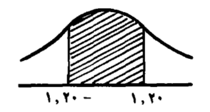
مثال (٢) : في دراسة عن خصائص ضغط الدم لسكان إحدى المدن وجد أن الوسط الحسابي لضغط الدم = ١٢١ وأن الانحراف المعياري = ١٢,٥ . يعطي شكل (٣) المدرج التكراري النسبي المناظر لهذه البيانات مقارناً بالمنحنى الطبيعي .

تنص القاعدة العملية على أن حوالي ٩٥٪ من المشاهدات تقع داخل بعد قدره انحرافين معياريين من الوسط الحسابي . ولتوضيح سبب ذلك يلاحظ أن نسبة السكان في شكل (٣) الذين يتراوح ضغط دمهم بين س ـ ٢ع ، س + ٢ع تمثل المساحة تحت المدرج التكراري المحصورة بين −٢ ، + ٢ وحدة معيارية ، وهي المساحة المظللة في الشكل . ولما كان الشكل العام للمدرج التكراري يقترب من شكل المنحنى الطبيعي ، فإن المساحة المظللة تساوي تقريباً المساحة المحصورة بين − ٢ ، + ٢ تحت المنحنى الطبيعي . وسنرى فيما بعد باستخدام الجداول الاحصائية المناسبة أن هذه المساحة تساوي ٠,٩٥ .

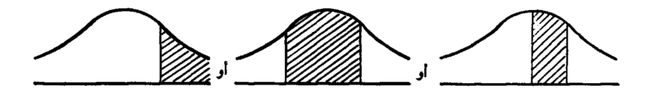
يتضح من هذين المثالين أن المنحنى الطبيعي يمكن أن يستخدم بشكل تقريبي لحساب المساحات المختلفة تحت المدرج التكراري النسبي . ويتطلب ذلك الاعتماد على الوسط الحسابي والانحراف المعياري لتحويل المشاهدات إلى وحدات معيارية ، ثم استخدام خصائص المنحنى الطبيعي لحساب قيم المساحات المطلوبة . ويؤكد ذلك مرة ثانية صلاحية الوسط الحسابي والانحراف المعياري في هذه الحالات لإعطاء وصف كامل لخصائص التوزيع التكراري .

٢ ـ إيجاد المساحات تحت المنحنى الطبيعي

يستخدم الجدول المعطى في نهاية هذا الباب لإيجاد المساحات تحت المنحنى الطبيعي . فمثلاً لإيجاد المساحة تحت المنحنى المحصورة بين ـ ١,٢٠، ١,٢٠ ، تقرأ قيمة م (ى) المناظرة لقيمة ى = ١,٢٠ في هذا الجدول . هذه القيمة تساوي ٠,٧٧ تقريباً . وبالتالي فإن ٧٧٪ من مساحة المنحنى الطبيعي تقع بين ـ ١,٢٠ ، ١,٢٠ كما يظهر في الشكل التالي .



قد تدعو الحاجة أيضاً إلى إيجاد مساحات على الشكل



وتوضح الأمثلة التالية كيفية إيجاد هذه المساحات .

مثال (٣) : أوجد قيمة المساحة بين صفر ، ١ تحت المنحنى الطبيعي .

الحل : تظهر المساحة المطلوبة في الشكل التالي :

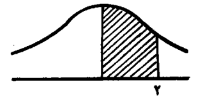


ولما كان الجدول يعطي المساحة من - ١ حتى + ١ وتساوي ٠,٦٨ تقريباً ، فإن المساحة من صفر إلى ١ تكون نصف هذه المساحة نتيجة تماثل

المنحنى الطبيعي ، أي أن المساحة المطلوبة تساوي $\frac{٠,٦٨}{٢}$ = ٠,٣٤

مثال (٤) : اوجد قيمة المساحة بين صفر ، ٢ تحت المنحنى الطبيعي .

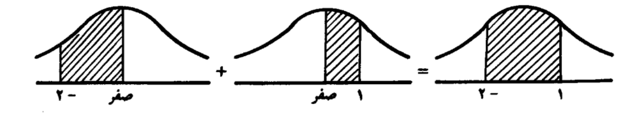
الحل :



تجدر الاشارة إلى أن هذه المساحة لا تساوي ضعف المساحة بين صفر ، ١ . وإنما يجب اتباع نفس الطريقة المتبعة في مثال (٣) ، حيث يلاحظ من الجدول أن المساحة المحصورة بين - ٢ ، ٢ تساوي تقريباً ٠,٩٥ . وبالتالي فإن المساحة المحصورة بين صفر ، ٢ تكون نصف هذا المقدار أي ٠,٤٧٥ نتيجة تماثل المنحنى الطبيعي .

مثال (٥) : اوجد قيمة المساحة بين - ٢ ، ١ تحت المنحنى الطبيعي .

الحل : يلاحظ أن المساحة بين - ٢ ، ١ هي مجموع المساحات بين



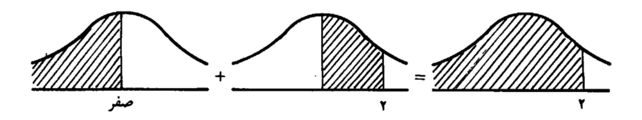
- ٢ ، صفر وبين صفر ، ١ ولما كانت المساحة بين - ٢ ، صفر تساوي المساحة بين صفر ، ٢ بسبب تماثل المنحنى ، فإن هذه المساحة تساوي ٠,٤٨ تقريباً . كذلك فإن المساحة بين صفر ، ١ تساوي ٠,٣٤ تقريباً وبالتالي تكون المساحة المطلوبة هي ٠,٤٨ + ٠,٣٤ =٠,٨٢ .

مثال (٦) : أوجد المساحة الى يمين القيمة ١ تحت المنحنى الطبيعي .

الحل : يعطي الجدول المساحة بين - ١ ، + ١ وهي ٠,٦٨ تقريباً . وبالتالي تكون المساحة خارج هذه الفترة مساوية ٠,٣٢ وبالتماثل تكون المساحة إلى يمين القيمة ١ مساوية نصف هذه القيمة أي ٠,١٦ .

مثال (٧) : أوجد المساحة إلى يسار القيمة ٢ تحت المنحنى الطبيعي .

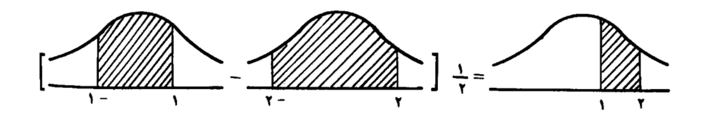
الحل : يلاحظ أن المساحة إلى يسار القيمة ٢ هي مجموع المساحات الى يسار القيمة صفر وبين صفر ، ٢ .



وتمثل المساحة الى يسار القيمة صفر نصف المساحة الكلية تحت المنحنى أي ٠,٥٠ بينما تمثل المساحة بين صفر ، ٢ حوالي ٠,٤٨ وبالتالي تكون المساحة المطلوبة هي ٠,٥٠ + ٠,٤٨ = ٠,٩٨ .

مثال (٨) : اوجد المساحة بين ١ ، ٢ تحت المنحنى الطبيعي .

الحل : تمثل المساحة المطلوبة نصف الفرق بين المساحة بين - ٢ ، ٢ والمساحة بين - ١ ، ١ .



ولما كانت المساحة بين - ٢ ، ٢ في الجدول تساوي تقريباً ٠,٩٥ والمساحة بين - ١ ، ١ تساوي تقريباً ٠,٦٨ فإن المساحة المطلوبة تساوي

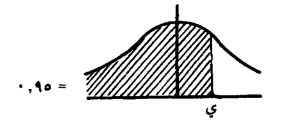
$\frac{1}{٢}$ (٠,٩٥ - ٠,٦٨) = ٠,١٤

ويلاحظ في جميع هذه الأمثلة عدم وجود قاعدة محددة لايجاد المساحات . كل ما هنالك هو محاولة وضع المساحة المطلوبة في شكل يسمح بإيجادها من الجدول .

يمكن أيضاً استخدام جدول مساحات المنحنى الطبيعي بشكل عكسي ، لتحديد القيمة التي تحصر إلى يسارها نسبة معلومة من مساحة المنحنى . وقد أشرنا إلى مثل هذه القيم بشكل عام في الأبواب السابقة ، ذلك أنه إذا كانت النسبة المعلومة = ٠,٢٥ فإن القيمة المناظرة هي الربيع الأول ، واذا كانت النسبة المعلومة = ٠,٧٥ فإن القيمة المناظرة هي الربيع الثالث ، واذا كانت النسبة المعلومة = ٠,٢٠ فإن القيمة المناظرة هي العشير الثاني وإذا كانت النسبة المعلومة = ٠,٩٥ فإن القيمة المناظرة هي المييء الخامس والتسعون ، وهكذا . ويوضح المثال التالي كيفية ايجاد هذه القيم .

مثال (٩) : احسب قيمة المييء الخامس والتسعون للمنحنى الطبيعي .

الحل : المطلوب حل المعادلة الآتية لتحديد قيمة ي :



إذا أريد استخدام جدول المساحات تحت المنحنى الطبيعي لحل هذه المعادلة ، يجب اعادة كتابتها بشكل يسمح باستخدام هذه الجداول ، وذلك لأن الجدول يعطي المساحة بين - ي ، ي . يلاحظ أن المساحة على يمين ي = ٠٥, وبالتالي يمكن من تماثل المنحنى ان نكتب :



وبالنظر في الجدول تحت م (ي) عند القيمة ٠,٩٠ نجد أي قيمة ي = ١,٦٥ تقريباً . وعلى ذلك فإن المبيىء الخامس والتسعين للمنحنى الطبيعي يساوي ١,٦٥ .

٣ ـ استخدام المنحنى الطبيعي كتقريب للبيانات

يستخدم المنحنى الطبيعي لحساب النسب المختلفة تحت المدرج التكراري لظاهرة بشكل تقريبي طالما أن الشكل العام لهذا المدرج التكراري يقترب من شكل المنحنى التكراري . ويتضح ذلك من الأمثلة التالية :

مثال (١٠) : كان الوسط الحسابي لبيانات ضغط الدم المستخدمة في مثال (٢) يساوي ١٢١ مم وكان الانحراف المعياري يساوي ١٢,٥ مم . استخدم المنحنى الطبيعي لتقدير نسبة السكان الذين يتراوح ضغط دمهم بين ٩٦مم ، ١٣٣,٥مم .

الحل : اذا أريد استخدام جدول المنحنى الطبيعي ، فلا بد من تحويل الفترة (٩٦ ، ١٣٣,٥) إلى وحدات معيارية .

الوحدات المعيارية المناظرة للقيمة ٩٦ = $\frac{٩٦ - ١٢١}{١٢,٥}$ = -٢

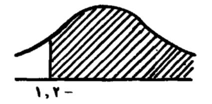
الوحدات المعيارية المناظرة للقيمة١٣٣,٥ = $\frac{١٣٣,٥ - ١٢١}{١٢,٥}$ = ١

وبالتالي تكون النسبة المحصورة تحت المدرج التكراري لضغط الدم بين ٩٦ ، ١٣٣,٥ مساوية تقريباً للنسبة المحصورة بين - ٢ ، ١تحت المنحنى الطبيعي . وتساوي هذه النسبة ٨٢٪ تقريباً ، باستخدام الجدول .



مثال (١١) : كان الوسط الحسابي لأوزان النساء المستخدمة في مثال (١) يساوي ٦٣كجم وكان الانحراف المعياري = ٢,٥كجم . استخدم المنحنى الطبيعي لحساب نسبة النساء اللائي يزيد وزنهن عن ٦٠ كجم.

الحل : هذه النسبة هي المساحة المحصورة تحت المدرج التكراري للأوزان إلى يمين القيمة ٦٠ . نحول هذه القيمة إلى وحدات معيارية ، حيث يلاحظ أن الوحدات المعيارية المناظرة ٦٠ـ٦٣ = ـ ١,٢ وبالتالي فإن النسبة المطلوبة هي المساحة تحت المنحنى الطبيعي إلى يمين القيمة



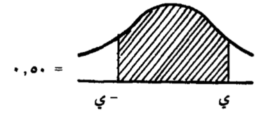
- ١,٢ . وباستخدام جدول المساحات تحت المنحنى الطبيعي نجد أن هذه النسبة تساوي ٠,٨٨ تقريباً .

مثال (١٢) : احسب قيمة كل من الربيع الأول والربيع الثالث لبيانات الأوزان المعطاة في مثال (١١) وذلك باستخدام المنحنى الطبيعي .

الحل : لايجاد الربيع الأول ، نبحث عن قيمة الوزن الذي يقل عنها قيم ٢٥٪ من المشاهدات . وتتمثل طريقة العمل في هذه الحالة في إيجاد الربيع الأول للمنحنى الطبيعي وتكون القيمة الناتجة معطاة بوحدات معيارية . تستخدم هذه القيمة بشكل عكسي للرجوع إلى قيمة المشاهدة المناظرة لهذه الوحدات المعيارية .



عند إيجاد قيمة الربيـع الأول للمنحنى الطبيعي فـإن ذلك يتـطلب حل المعادلة الآتية لتحديد قيمة - ي ، حيث يلاحظ أن قيمة الربيع الأول في هذه الحالة تكون سالبة لأن أي قيمة مناظرة لنسبة تقل عن ٥٠ وتقع بالتعريف إلى يسار الصفر . يمكن اعادة كتابة المعادلة كما يلي :



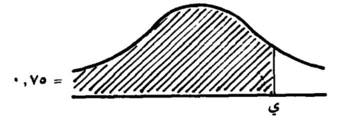
ومنه نستنتج أن ـ ي = ـ ٠,٦٨ ويتم تحديد قيمة المشاهدة المناظرة بحل المعادلة :

$$- ٠,٦٨ = \frac{\text{قيمة المشاهدة ـ قيمة الوسط الحسابي}}{\text{قيمة الانحراف المعياري}}$$

$$- ,٦٨ = \frac{\text{المشاهدة ـ ٦٣}}{٢,٥} \text{ ،}$$

ومنه نستنتج أن قيمة الربيع الأول = ٦١,٣كجم .

بنفس الـطريقة ، يتحـدد الربيـع الأعلى على المنحنى الـطبيعي بحـل المعادلة :



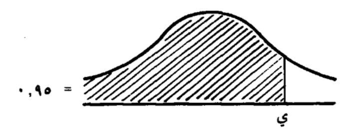
ومنـه نستنتج أن ي = ٠,٦٨ ويتحـدد الـربيـع الأعلى لـلأوزان بحـل المعادلة :

$$٠,٦٨ = \frac{\text{المشاهدة - ٦٣}}{٢,٥} \text{ ،}$$

أي أن قيمة الربيع الأعلى = ٦٤,٧كجم .

مثال (١٣) : كان توزيع درجات الطلبة في امتحان القبول للجامعة قريب من شكل المنحنى الطبيعي بوسط حسابي = ٦٥٠درجة وانحراف معياري = ٦٠ درجة . احسب قيمة المييء الخامس والتسعين لهذه البيانات .

الحل : نبدأ بإيجاد قيمة المييء الخامس والتسعين للمنحنى الطبيعي ، وذلك بحل المعادلة الآتية لتحديد قيمة ي .



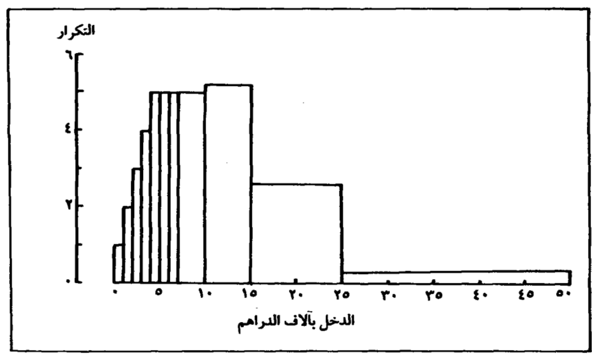
وتكون قيمة ي من الجدول مساوية ١,٦٥ تقريباً .
لتحديد قيمة المييء الخامس والتسعين للدرجات ، تحل المعادلة :

$$١,٦٥ = \frac{\text{المييء الخامس والتسعين} - ٦٥٠}{٦٠}$$

ومنه تكون قيمة المييء الخامس والتسعين مساوية ٧٤٩ درجة .

سبقت الاشارة إلى أن الكثير من المدرجات التكرارية تتبع الشكل العام للمنحنى الطبيعي . وتوضح الأمثلة السابقة أن الوسط الحسابي والانحراف المعياري تكون مقاييس جيدة لتلخيص نمط الاختلاف في مثل هذه البيانات . إلا أنه يجب أن نذكر أن هناك مدرجات تكرارية أخرى لا يكون شكلها قريب من المنحنى الطبيعي . ويكون الوسط الحسابي والانحراف المعياري في مثل هذه الحالات مقاييس غير جيدة لوصف التوزيع . فمثلاً ، يعطي شكل (٤) التوزيع التكراري للدخل الشهري للأسر في دولة ما . وقد كان الوسط الحسابي لهذه البيانات يساوي ١٤ ألف درهم تقريباً ، كما كان الانحراف المعياري مساوياً ١٠ آلاف درهم تقريباً . ويظهر تطبيق المنحنى الطبيعي على

هـذه البيانـات أن نسبة الأسـر التي تحصل على دخـل سـالب (أي أقـل من الصفر) تناظـر المساحـة تحت المنحنى الطبيعي إلى يسـار القيمة المعيـارية -١,٤ (أي $\frac{\text{صفر} - ١٤٠٠٠}{١٠٠٠٠}$) ، وهي مساحة تساوي ٠٨, تقريباً. وهذه نتيجة غير معقولة اذ لا يوجد دخل سالب ، ويرجع السبب في ذلك إلى أن المدرج التكراري في شكل (٤) لا يشبه المنحنى الطبيعي ، اذ أنه مدرج غير متماثل وملتو بشكل واضح إلى اليمين .



شكل (٤) : المدرج التكراري لتوزيع الدخول يبعد شكله
عن شكل التوزيع الطبيعي

لا ينصح في مثل هذه الحالات بالاعتماد على الوسط الحسابي والانحراف المعياري لوصف التوزيع وإنما يفضل حساب مقاييس ترتيبية مثل الربيع الأول والوسيط والربيع الثالث والعشيرات والميئيات المختلفة .

٤ ـ دراسة أخطاء القياس

يتوقع نظرياً عند قياس قيمة متغير متصل لمفردة ما ، الحصول على نفس النتيجـة في كل مـرة تتكـرر فيهـا عمليـة القيـاس لنفس المفـردة تحت نفس

الظروف . ولكن ذلك لا يحدث عملياً . فمثلاً ، اذا استخدمت أداة قياس حساسة لقياس طول مسافة معينة عدداً من المرات المتتالية ، فإن نتائج عملية القياس تختلف من مرة إلى أخرى . وتثير هذه الحقيقة تساؤلاً حول مفهوم درجة دقة أداة القياس ذاتها . وتجدر الاشارة إلى أن أخطاء القياس هذه تحدث فعلاً على الرغم مما تقوم به الأجهزة المسئولة عن المقاييس والموازين في الدول المختلفة بمراقبة وضبط أدوات القياس بشكل دوري .

وتعتمد أهمية دراسة درجة دقة أدوات القياس على مستوى الدقة المطلوب في النتائج . اذ تكون مثل هذه الدراسة ضرورية في المواقف التي تتطلب درجة دقة عالية في القياس مثل التطبيقات المعملية الهندسية المختلفة . ويجب على الباحثين ، في هذه الحالات ، دراسة أنماط أخطاء القياس واقتراح الأساليب الملائمة لعلاجها .

وتمثل دراسة أنماط الخطأ في عمليات القياس إحدى التطبيقات الاحصائية المفيدة . وتعتمد مثل هذه الدراسة على استخدام اداة القياس للحصول على قراءات متكررة لنفس المفردة ، تحت نفس الظروف ، ثم تحليل الاختلافات المشاهدة بين هذه القراءات . ويتم عادة ، في هذا الصدد حساب قيمة الانحراف المعياري ثم افتراض أن نتائج عملية القياس تتبع المنحنى الطبيعي . وتتضح هذه الأمور بمراجعة المثال التالي الذي ورد في كتاب Freedman et al. (انظر قائمة المراجع المختارة في نهاية الكتاب) .

مثال (١٤) : تقوم ادارة المقاييس والموازين في دولة ما بمراقبة وضبط أجهزة الوزن المختلفة في الدولة بشكل دوري . وتعتمد الادارة في ذلك على وحدات عيارية تحتفظ بها . فمثلاً هناك ثقل يتفق على أن وزنه يساوي ١ كجم يستخدم كعيار لمراجعة أوزان الكيلوجرامات الأخرى ، وهكذا .

وتقوم الادارة بوزن هذه الأوزان العيارية دورياً (مرة كل أسبوع) للتأكد من درجة دقتها ، وذلك من خلال دراسة الاختلافات في الوزن من أسبوع لآخر . يعطي المثال الحالي نتائج هذه القياسات التي أجريت على الثقل المعياري المساوي ١٠ جم أسبوعياً لمدة مائة أسبوع في هذه الدولة . وتجدر

الاشارة إلى أن هذه القياسات تتم في نفس الغرفة وباستخدام نفس الاجهـزة وبالتحكم بقدر الامكان في العوامل التي يمكن أن تؤثر في نتيجة القياس مثل درجات الحرارة أو ضغط الهواء ، . . . الخ .

كانت القراءات الخمس الأولى هي :

٩,٩٩٩٥٩١جـم ، ٩,٩٩٩٦٠٠جـم ، ٩,٩٩٩٥٩٤ جم ، ٩,٩٩٩٦٠١جم ، ٩,٩٩٩٥٩٨جم.

وتوضح هذه البيانات خصائص عمليات القياس الدقيقة ، اذ يلاحظ أن المشـاهدات المتكـررة تختلف فيما بينهـا فقط في الأرقام العشريـة الثـلاث الأخيرة . ولما كانت هذه الاختلافات لا يمكن ارجاعها لسبب معروف فإنهـا تسمى أخطاء الصدفة .

جدول (١)

عدد الميكروجرامات الناقصة عند وزن
الثقل المعياري ١٠جم مائة مرة تحت نفس الظروف

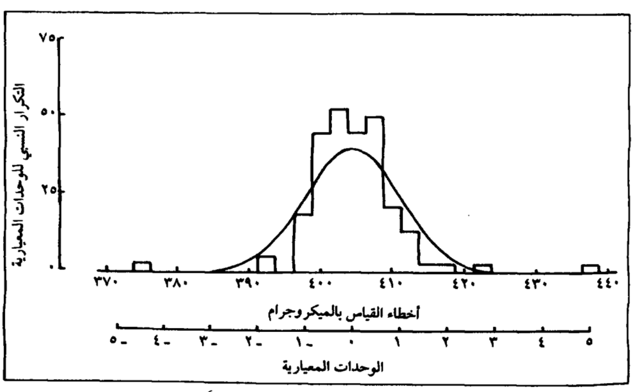
٤٠٦	٤٠٢	٣٩٨	٤٠١	٤٠٥	٣٩٩	٤٠٧	٤٠٣	٤٠٧	٤٠١
٤٠٠	٤٠٦	٤٠٣	٣٩٩	٤٠٢	٤٠٢	٤٠١	٤٠٠	٤٢٣	٤٠٣
٤٠٦	٤٠١	٤٠٧	٤٠٠	٤٠٨	٣٩٩	٣٩٩	٤١٠	٤٠٦	٤٠٢
٣٩٩	٤٠٣	٤٠٢	٤٠١	٣٩٩	٣٩٧	٤٠١	٤٠١	٤٠٦	٤٠٥
٤٠٥	٤١٣	٤٠٧	٤٠١	٤٠٦	٤١٢	٤٠١	٤٠١	٤٠٤	٤٠٩
٤٠٩	٤٠٩	٤٠٥	٤٠٢	٤٠٤	٤٠٦	٤٠٤	٤١٢	٤٠١	٤٠٦
٣٩٩	٤٠٤	٤١١	٤٠٤	٤٠٣	٤٠٩	٤٠٨	٣٩٣	٤٠١	٣٩٨
٤٠٢	٤٠٢	٤١٠	٤٠٥	٤٠٨	٤٠٠	٤٠٦	٤٣٧	٤٠٧	٤٠٦
٤٠٧	٤٠٤	٤١٠	٣٩٢	٤٠٤	٤٠٨	٤٠٨	٤١٨	٤١٢	٤٠٣
٤٠٦	٤٠٦	٤١٠	٤٠٧	٤٠٧	٤٠٤	٤٠٦	٤١٥	٣٧٥	٤٠٤

لاحظ الباحثون أن هذه البيانات تدل على أن الثقل المعياري يقل وزنه دائماً عن ١٠جم ، ولذلك تقرر الاكتفاء بتسجيل حجم هذا الخطأ للمشاهدات

المختلفــة . فمثـلاً يكــون حجم الخـطأ في المشـاهـدة الأولى مســاويــاً ٤٠٩٠٠٠, جم أو ٤٠٩ ميكروجرام (الميكروجرام يساوي واحد على المليون من الجرام) . وتظهر جميع المشاهدات المائة في جدول (١) .

يلاحظ أن الوسط الحسابي لهذه الأخطاء = ٤٠٥ ميكروجرام تقريباً وأن الانحراف المعياري يساوي ٦ ميكروجرام تقريباً . لاحظ أن هذه النتائج تدل على الدقة العالية لهذا الوزن المعياري لأن ٤٠٥ ميكروجرام لا تتعدى وزن ذرة من الملح . وتفسر هذه النتائج بـأن الوزن الحقيقي للثقـل المعياري يقل عن ١٠جم بمقـدار قريب من ٤٠٥ ميكـروجرام ، وأن تقـديراً للقيمـة المتوسـطة لاختلاف كل قـراءة عن هذا الـوزن الحقيقي يساوي ٦ ميكـروجرام تقـريباً . ويمكن اعتبـار القيمة ٤٠٥ تقـديراً لمقـدار التحيز في وزن الثقـل المعيـاري واعتبار القيمة ٦ ميكروجرام مقياساً للتفاوت في قيمة هذا التحيز من مشـاهدة لأخرى . وتعتمد دقة هذه التفسيـرات على الشكل العـام للمدرج التكـراري للبيانات ودرجة اقترابه من المنحنى الطبيعي .

يعـطي شكل (٥) المـدرج التكراري المنـاظر لهـذه البيـانـات مقـارنـاً

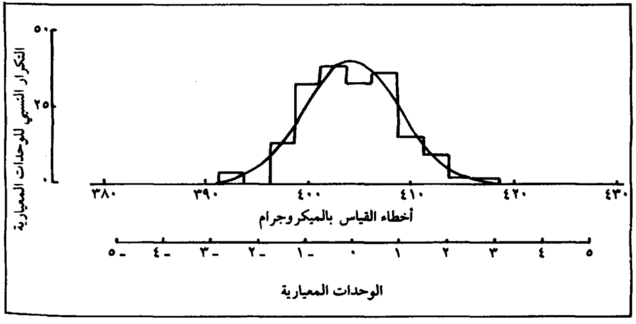


شكل (٥) : المدرج التكراري لأخطاء القياس مقارناً بالمنحنى الطبيعي

بالمنحنى الطبيعي ، حيث يلاحظ وجود اختلافات أساسية بينهما . يلاحظ وجود قيم شاذة أو متطرفة على المدرج التكراري وهي ٣٧٥ ، ٤٢٣ ، ٤٣٧ ، اذ تقع هذه المشاهدات خارج نطاق المنحنى التكراري .

إذا استبعدت هذه المشاهدات المتطرفة من البيانات ، فإن قيمة الوسط الحسابي للمشاهدات الباقية = ٤٠٤ ميكروجرام والانحراف المعياري = ٤ ميكروجرام (لاحظ التأثر الواضح للانحراف المعياري بهذه القيم المتطرفة) . يعطي شكل (٦) المدرج التكراري المناظر مقارناً بالمنحنى الطبيعي . ويلاحظ أن شكل المدرج يقترب من المنحنى الطبيعي في هذه الحالة.

وتجدر الاشارة إلى أن الحصول على عدد قليل من المشاهدات الشاذة أمر يحدث دائماً عند اجراء القياسات المختلفة . وقد جرت العادة على اهمال هذه المشاهدات واستبعادها نظراً لعدم القدرة على تفسير أسباب وقوعها . ولا يدخل ضمن ذلك ، بالطبع ، المشاهدات الشاذة التي تنشأ بسبب اختلال شروط اجراء التجربة وجمع البيانات ، اذ يجب دراسة مثل هذه المشاهدات على حدة وذلك بهدف التعرف على أساليب تفادي حدوثها في التجارب المستقبلية .



شكل (٦) : المدرج التكراري لأخطاء القياس
بعد استبعاد القيم المتطرفة ومقارنته بالمنحنى الطبيعي

ونختتم هذا الباب بملاحظة أن أهمية المنحنى الطبيعي في الدراسات الاحصائية لا تقتصر على استخدامات هذا المنحنى لتمثيل التوزيع التكراري للمتغيرات المختلفة بشكل تقريبي . ذلك أن المنحنى الطبيعي يلعب دوراً اكثر أهمية في مجالات الاستنتاج الاحصائي . وهناك عدد من النظريات الرياضية التي توضح كيف أن نمط اختلاف قيم احصاءات العينة من عينة لأخرى يمكن تمثيله بالمنحنى الطبيعي . ويتخذ ذلك كأساس للاستفادة بهذه الاحصاءات عند تقدير معالم المجتمع . وسوف نناقش هذا الأمر بالتفصيل عند دراسة أساليب الاستنتاج الاحصائي .

تمرينات

١ ـ في دراسة عن أنماط الدرجات التي يحصل عليها طلبة الثانوية العامة في أحد البلدان ، لوحظ أن مجاميع الطلبة تزيد من عام لآخر . فمثلاً كان متوسط مجموع درجات الطالب في عام ١٩٨٠ يساوي ٤٤٥ وفي عام ١٩٨٥ يساوي ٤٦٥ . اذا علم أن قيمة الانحراف المعياري ثابتة في هذين العامين وتساوي ١٠٠ وأن الشكل العام للمدرج التكراري قريب من المنحنى الطبيعي فاوجد :

(أ) نسبة الطلبة الذين يحصلون على مجموع يزيد عن ٦٠٠ في عام ١٩٨٠ .

(ب) نسبة الطلبة الذين يحصلون على مجموع يزيد عن ٦٠٠ في عام ١٩٨٥ .

٢ ـ استخدم جدول المساحات تحت المنحنى الطبيعي لايجاد المساحات التالية :

(أ) المساحة تحت المنحنى بين صفر ، ١,٦٥ .

(ب) المساحة تحت المنحنى بين - ١,٣٠ ، صفر .

(حـ) المساحة تحت المنحنى الى يمين القيمة ١,٩٥ .

(د) المساحة تحت المنحنى إلى يسار القيمة ٢,٠٠ .

٣ ـ اذا كان معلوماً أن الوسط الحسابي لقياسات درجة الذكاء لأطفال المدارس يساوي ١٠٠ وأن الانحراف المعياري يساوي ٢٠ . استخدم المنحنى الطبيعي لحساب النسب الآتية :

(أ) نسبة الأطفال الذين يقل قياس درجة ذكائهم عن ١٣٠ .

(ب) نسبة الأطفال الذين يتراوح قياس درجة ذكائهم بين ١١٠ ، ١٣٠ .

(حـ) نسبة الاطفال الذين يتراوح قياس درجة ذكائهم بين ٨٠ ، ١١٠ .

(د) نسبة الأطفال الذين يقل قياس درجة ذكائهم عن ٦١ أو يزيد عن ١٣٩ .

(هـ) نسبة الأطفال الذين يزيد قياس درجة ذكائهم عن ١٣٣ .

٤ ـ استخدم المعلومات المعطاة في تمرين (٣) لحساب الآتي :

(أ) قيمة الربيع الأول لقياسات درجة الذكاء .

(ب) قيمة الربيع الثالث لقياسات درجة الذكاء .

(حـ) قيمة العشير السادس لقياسات درجة الذكاء .

(د) قيمة المييء التسعون لقياسات درجة الذكاء .

٥ ـ اذا كان معلوماً أن طول محيط الرقبة للذكور البالغين في بلد ما له توزيع قريب من المنحنى الطبيعي بوسط حسابي = $\frac{١}{٢}$ ١٥ بوصة وانحراف معياري = $\frac{٣}{٤}$ بوصة . ينتج أحد مصانع القمصان ثلاث مقاسات مختلفة هي صغير (ويصلح لمن يقل طول محيط رقبتهم عن ١٥ بوصة) ومتوسط (لمن يتراوح طول محيط رقبتهم بين ١٥ بوصة ، $\frac{١}{٢}$ ١٦ بوصة) وكبير (لمن يزيد طول محيط رقبتهم عن $\frac{١}{٢}$ ١٦ بوصة) . احسب نسبة كل من هذه المقاسات في انتاج المصنع .

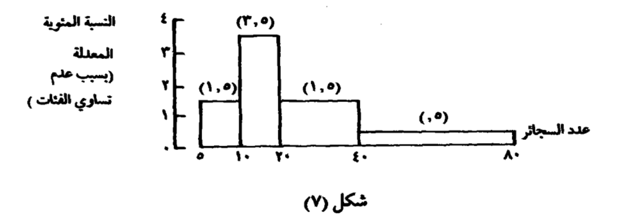
٦ ـ بلغ متوسط درجة الطالب في أحد امتحانات الاحصاء ٥٥ درجة وكان الانحراف المعياري يساوي ١٨ درجة . يرغب استاذ المساق في اعطاء تقدير ممتاز لنسبة ١٥٪ من الطلبة وتقدير جيد جداً لنسبة ٢٠٪ من الطلبة وتقدير جيد لنسبة ٤٠٪ من الطلبة وتقدير مقبول لنسبة ٢٠٪ من الطلبة وتقدير راسب لنسبة ٥٪ من الطلبة . استخدم المنحنى الطبيعي لتحديد المدى المناظر لدرجات كل تقدير .

٧ ـ هناك مائة مشاهدة عن متغير ما . حولت هذه المشاهدات إلى وحدات معيارية ، باستخدام برنامج للحساب على الحاسب الآلي . فيما يلي النتائج التي اعطيت للمشاهدات العشر الأولى :

-٢,٦ ، ٣,٥ ، ٢,١ ، -١٣,٠ ، ٤,٢ ، -١,٥ ، -٢,٧ ، -٣,١١ ،

١,٨ ، ٦,٣ هل تعتقد أن هناك خطأ ما في حساب هذه الوحدات المعيارية ؟ وضح سبب اجابتك .

٨ ـ يعطي شكل (٧) المدرج التكراري المناظر لعدد السجائر التي يدخنها المدخنون في دراسة معينة . كان الوسط الحسابي لعدد السجائر في هذا التوزيع يساوي ٢٧ والانحراف المعياري يساوي ٢٠ . استخدم المنحنى الطبيعي لتقدير نسبة الأشخاص الذين يدخنون سيجارتين على الأكثر . احسب هذه النسبة مباشرة من المدرج التكراري وقارن بين النتيجتين موضحاً سبب الاختلاف بينهما .



شكل (٧)

٩ ـ اذكر ما إذا كان كل من العبارات الآتية صحيح أم خطأ ، مع شرح سبب الاجابة .

(أ) «يقع نصف المشاهدات دائماً إلى يسار المتوسط» .

(ب) «اذا كان هناك مجموعتين من المشاهدات وكان الوسط الحسابي والانحراف المعياري لمشاهدات كل من المجموعتين يساوي ٥٠، ١٠ على الترتيب ، فإن نسبة المشاهدات التي تقع بين ٤٠، ٦٠ لا بد وأن تكون متساوية في المجموعتين .

١٠ ـ فيما يلي الأجر اليومي في عينة من ٢٣ عاملاً ، حيث يلاحظ أن الوسط الحسابي = ٥٠ درهم ، وأن الانحراف المعياري = ١٠ دراهم :

٣٩ ٤١ ٤٧ ٥٨ ٦٥ ٣٧ ٣٧ ٤٩ ٥٦ ٥٩ ٦٢ ٣٦ ٤٨
٥٢ ٦٤ ٢٩ ٤٤ ٤٧ ٤٩ ٥٢ ٥٣ ٥٤ ٧٢ ٥٠ ٥٠

(أ) استخدم المنحنى الطبيعي لتقدير عدد المشاهدات التي تقع داخل بعد قدره ١,٢٥ انحراف معياري من الوسط الحسابي .
(ب) احسب عدد هذه المشاهدات مباشرة في البيانات الأصلية .
(جـ) قارن بين النتيجة في (أ) والنتيجة في (ب) .

١١ ـ أرسل أحد المعامل ثقلاً لادارة المقاييس والموازين طالباً تحديد وزنه . قامت الادارة بتحديد الوزن فكان ١٦,٠٠٧ أوقية . اذا أرسل المعمل هذا الثقل مرة ثانية للادارة وقامت الادارة بوزنه فهل تعتقد أن هذا الوزن سيكون (أ) أم (ب) أم (جـ) ؟ ولماذا ؟
(أ) ١٦,٠٠٧ تماماً .
(ب) ١٦,٠٠٧ ± ٠,٠٠١
(حـ) ١٦,٠٠٧ ± ٠,٠١

١٢ ـ ما معنى أن تكون قيمة الانحراف المعياري لمجموعة من المشاهدات تساوي الصفر .

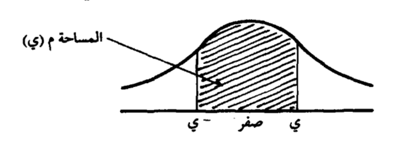
١٣ ـ كان الوسط الحسابي لدرجات الطلبة في الامتحان النهائي للإحصاء يساوي ٥٠ والانحراف المعياري = ٢٠ وكان المدرج التكراري المناظر قريب الشبه بالمنحنى الطبيعي . اختير أحد الطلبة عشوائياً وطلب منك تخمين الدرجة التي حصل عليها هذا الطالب . ما هو احتمال أن الدرجة التي تخمنها لن تختلف عن الدرجة الفعلية للطالب بأكثر من ١٠ درجات .

١٤ ـ طلب من كل طالب في عينة من ١٩ طالباً قياس سمك إحدى القطع الخشبية لأقرب جزء من الألف من البوصة . قام كل طالب بقياس هذه القطعة الخشبية مرتين ، وفيما يلي نتائج هذه القياسات :

رقم الطالب	نتيجة القياس في المرة الأولى	نتيجة القياس في المرة الثانية
١	١,٣١٧	١,٣٢٠
٢	١٣,٢٦	١٣,٢٥
٣	١,٣١٦	١,٣٣٥
٤	١,٣١٦	١,٣٢٨
٥	١,٣١٨	١,٣٢٤
٦	١,٣٢٩	١,٣٢٦
٧	١,٣٣٢	١,٣٣٤
٨	١,٣٤٢	١,٣٢٨
٩	١,٣٣٧	١,٣٤٢
١٠	١٣,٢٦	١٣,٢٥
١١	١,٣٣٣	١,٣٣٤
١٢	١,٣١٥	١,٣١٧
١٣	١,٣١٦	١,٣١٨
١٤	١,٣٢١	١,٣١٩
١٥	١,٣٣٧	١,٣٤٣
١٦	١,٣٤٩	١,٣٣٦
١٧	١,٣٢٠	١,٣٣٦
١٨	١,٣٤٢	١,٣٤٠
١٩	١,٣١٧	١,٣١٨

هل تدل هذه البيانات على أن كل طالب قد قام بقياس القطعة الخشبية بمعزل عن الطلبة الآخرين أم هناك ما يدل على تعاون الطلبة فيما بينهم للحصول على هذه القياسات . اشرح سبب إجابتك .

جدول المساحات تحت المنحنى الطبيعي



ي	م (ي)	ي	م (ي)	ي	م (ي)	ي	م (ي)
,٠٠	,٠٠٠٠	١,٠٠	,٦٨٢٧	٢,٠٠	,٩٥٤٥	٣,٠٠	,٩٩٧٣
,٠٥	,٠٣٩٩	١,٠٥	,٧٠٦٣	٢,٠٥	,٩٥٩٦	٣,٠٥	,٩٩٧٧
,١٠	,٠٧٩٧	١,١٠	,٧٢٨٧	٢,١٠	,٩٦٤٣	٣,١٠	,٩٩٨١
,١٥	,١١٩٢	١,١٥	,٧٤٩٩	٢,١٥	,٩٦٨٤	٣,١٥	,٩٩٨٤
,٢٠	,١٥٨٥	١,٢٠	,٧٦٩٩	٢,٢٠	,٩٧٢٢	٣,٢٠	,٩٩٨٦
,٢٥	,١٩٧٤	١,٢٥	,٧٨٨٧	٢,٢٥	,٩٧٥٦	٣,٢٥	,٩٩٨٩
,٣٠	,٢٣٥٨	١,٣٠	,٨٠٦٤	٢,٣٠	,٩٧٨٦	٣,٣٠	,٩٩٩٠
,٣٥	,٢٧٣٧	١,٣٥	,٨٢٣٠	٢,٣٥	,٩٨١٢	٣,٣٥	,٩٩٩٢
,٤٠	,٣١٠٨	١,٤٠	,٨٣٨٥	٢,٤٠	,٩٨٣٦	٣,٤٠	,٩٩٩٣
,٤٥	,٣٤٧٣	١,٤٥	,٨٥٢٩	٢,٤٥	,٩٨٥٨	٣,٤٥	,٩٩٩٤
,٥٠	,٣٨٢٩	١,٥٠	,٨٦٦٤	٢,٥٠	,٩٨٧٦	٣,٥٠	,٩٩٩٥
,٥٥	,٤١٧٧	١,٥٥	,٨٧٨٩	٢,٥٥	,٩٨٩٢	٣,٥٥	,٩٩٩٦
,٦٠	,٤٥١٥	١,٦٠	,٨٩٠٤	٢,٦٠	,٩٩٠٧	٣,٦٠	,٩٩٩٧
,٦٥	,٤٨٤٣	١,٦٥	,٩٠١١	٢,٦٥	,٩٩٢٠	٣,٦٥	,٩٩٩٧
,٧٠	,٥١٦١	١,٧٠	,٩١٠٩	٢,٧٠	,٩٩٣١	٣,٧٠	,٩٩٩٨
,٧٥	,٥٤٦٧	١,٧٥	,٩١٩٩	٢,٧٥	,٩٩٤٠	٣,٧٥	,٩٩٩٨
,٨٠	,٥٧٦٣	١,٨٠	,٩٢٨١	٢,٨٠	,٩٩٤٩	٣,٨٠	,٩٩٩٩
,٨٥	,٦٠٤٧	١,٨٥	,٩٣٥٧	٢,٨٥	,٩٩٥٦	٣,٨٥	,٩٩٩٩
,٩٠	,٦٣١٩	١,٩٠	,٩٤٢٦	٢,٩٠	,٩٩٦٣	٣,٩٠	,٩٩٩٩
,٩٥	,٦٥٧٩	١,٩٥	,٩٤٨٨	٢,٩٥	,٩٩٦٨	٣,٩٥	,٩٩٩٩



الباب التاسع

الارتباط بين متغيرين

ناقشنا في الأبواب السابقة كيفية وصف نمط الاختلاف في مجموعة بيانات تتألف من مشاهدات عن متغير إحصائي واحد . ولا يهتم الاحصائيون بتحليل مشاهدات عن متغير واحد فقط ، بل يتطلب الأمر في معظم التطبيقات الاحصائية دراسة مشاهدات عن عدد من المتغيرات الاحصائية في آن واحد ، ووصف النمط المشاهد للعلاقات بين هذه المتغيرات . فمثلاً ، قد يهتم باحث زراعي عند تحليل نمط نمو نوع معين من الأشجار بدراسة العلاقة بين عمر الشجرة وطولها ، أو قد يتطلب بحث اجتماعي عن أنماط الزواج دراسة العلاقة بين عمر الزوج وعمر الزوجة ، أو قد يكون من المهم عند دراسة أسعار السيارات المستعملة ، أن توصف العلاقة بين سعر السيارة وعمرها ، أو قد يقوم المسؤلون عن التعليم العالي في الدولة بدراسة العلاقة بين مجموع درجات الطالب في الثانوية العامة ومدى تقدمه في دراسته الجامعية ، أو قد يكون الهدف هو دراسة العلاقة بين رأي الشخص في النظام المقترح للمواصلات العامة بالمدينة ومستوى دخله ، . . . وهكذا .

تسمى مجموعة البيانات التي تتألف من مشاهدات عن عدة متغيرات إحصائية بمجموعة بيانات متعددة المتغيرات Multivariate Data Set . ولعل أكثر هذه المجموعات أهمية في الاستخدامات الاحصائية المختلفة هي تلك التي تحتوي على بيانات عن متغيرين فقط . وفي هذه الحالة ، يرمز لأحد هذين المتغيرين بالرمز س بينما يرمز للمتغير الآخر ص . فمثلاً قد يكون س

هو عمر الشجرة ويكون ص هو طولها ، أو قد يكون س هو عمر السيارة ويكون ص سعرها أو قد يكون س هو دخل الشخص ويكون ص هو رأيه في النظام المقترح للمواصلات العامة ، . . . وهكذا . وتسمى مجموعة البيانات في هذه الحالة بمجموعة بيانات مزدوجة . ويلاحظ أن المشاهدة المأخوذة عن كل مفردة تتألف من قراءتين احداها للمتغير س والأخرى للمتغير ص . فمثلاً اذا كان هناك بيانات عن عمر السيارة (س) وسعر السيارة (ص) لمجموعة من ١٠ سيارات ، فإن معنى ذلك أن هناك قراءتين لكل سيارة تمثل احداها عمر السيارة وتمثل الأخرى سعرها. وتأخذ البيانات في هذه الحالة شكلاً مماثلاً للشكل التالي :

رقم السيارة	عمر السيارة (س)	سعر السيارة (ص)
١	٣	١٢٠٠٠
٢	٥	٦٠٠٠
.	.	.
.	.	.
.	.	.
١٠	٢	١٤٠٠٠

وترجع أهمية جمع وتحليل مجموعات البيانات المزدوجة إلى أن دراسة واستنتاج العلاقات بين المتغيرات المختلفة تمثل ركناً أساسياً في جميع مجالات العلم والمعرفة . ويمكن في هذا الصدد التمييز بين العلاقات التي تنشأ في مجالات العلوم الطبيعية وبين تلك التي تنشأ في مجالات العلوم الاجتماعية وعلوم الحياة . ذلك أن العلاقات التي تنشأ في مجالات العلوم الطبيعية تكون عادة علاقات تامة يحكمها قانون جبري . مثال ذلك العلاقة بين طول ضلع المربع ومساحته ، أو العلاقة بين سرعة سقوط جسم ما والجاذبية الأرضية أو العلاقة بين درجة الحرارة وضغط الغاز في اناء مغلق ، . . . وهكذا . وقد اهتم الباحثون منذ القدم بعملية استنتاج هذه القوانين . وقد يتم

ذلك بناءاً على الاعتبارات النظرية التي تحكم العلاقة أو بناء على نتائج تجارب مصممة خصيصاً لهذا الغرض . ولا تستخدم الأساليب الاحصائية بشكل أساسي في هذه العملية ، اللهم إلا في حدود ضيقة لوصف الأخطاء التي قد تحدث أثناء عمليات القياس المختلفة .

ويختلف الحال في مجالات العلوم الاجتماعية وعلوم الحياة . ذلك أن العلاقات بين المتغيرات المختلفة في هذه المجالات لا ترقى إلى مستوى القوانين الجبرية . مثال ذلك العلاقة بين عمر الزوج وعمر زوجته ، أو العلاقة بين مستوى تعليم الشخص ودخله ، أو العلاقة بين عمر الشجرة وطولها ، . . . وهكذا . ويتم الاعتماد بشكل أساسي على أساليب إحصائية خاصة لدراسة وتحليل مثل هذه العلاقات . ويهدف استخدام هذه الأساليب الى قياس قوة العلاقة بين المتغيرين ، ثم وصف هذه العلاقة بشكل تقريبي يسمح بالتعرف على متوسط قيمة التغير الذي يحدث في أحد المتغيرين ، المناظرة لزيادة قيمة المتغير الآخر بمقدار وحدة قياس واحدة .

يهتم هذا الباب والباب التالي بمناقشة هذه الأساليب الاحصائية وتوضيح استخداماتها المختلفة . وسينصب الاهتمام بشكل خاص على طرق الحساب وكيفية تفسير النتائج ، مع بيان أوجه القصور المختلفة في هذه الأساليب .

١ - بعض المفاهيم الأساسية

تتطلب الدراسة الاحصائية للعلاقة بين متغيرين استخدام أساليب خاصة لوصف وتحليل مجموعة بيانات مزدوجة عن هذين المتغيرين . وفي هذا الصدد ، هناك عدة مفاهيم أساسية ينبغي مناقشتها قبل البدء في عرض هذه الأساليب . ويشمل ذلك مفهوم التنبؤ ومفهوم الارتباط ومفهوم السببية .

يستخدم التنبؤ في الحياة اليومية عادة للدلالة على سلوك حدث مستقبلي . مثال ذلك التنبؤ بالطقس أو التنبؤ بعدد السكان في المستقبل أو التنبؤ بنتائج مباريات الأسبوع القادم . ولا يتفق ذلك مع المفهوم الاحصائي

للتنبؤ ، والذي يشمل أيضاً سلوك أحداث حاضرة أو أحداث ماضية . فمثلاً قد يراد التنبؤ بتكلفة الوقود المستخدم في تسيير سيارة ما خلال العام الماضي . في هذه الحالة يتم الاعتماد على معلومات متاحة مثل المسافات التي قطعتها هذه السيارة خلال العام ومعدل استهلاكها للوقود ، وذلك للحصول على التنبؤ المطلوب . ويتمثل مفهوم التنبؤ الاحصائي في محاولة تقدير قيمة متغير ما باستخدام معلومات متاحة عن متغيرات أخرى . وبعبارة أخرى ، إذا علمت قيمة متغير ما س ، فكيف يمكن تقدير أو التنبؤ بالقيمة المناظرة لمتغير آخر ص . فمثلاً ، كيف يمكن التنبؤ بقيمة المنفق على الطعام والشراب شهرياً لأسرة عدد أفرادها يساوي ستة أفراد ؟ (هنا المتغير س هو عدد أفراد الأسرة والمتغير ص هو قيمة المنفق على الطعام والشراب) . ويتم ذلك بجمع بيانات مزدوجة عن عدد أفراد الأسرة وقيمة المنفق على الطعام والشراب لمجموعة من أسر المجتمع محل الدراسة . وتفيد هذه البيانات في توضيح شكل العلاقة بين هذين المتغيرين ، التي يعتمد عليها للحصول على قيمة التنبؤ المطلوبة .

جرت العادة على أن يسمى المتغير الذي تستخدم قيمته كأساس للتنبؤ بالمتغير المستقل ويرمز له بالرمز س ، وأن يسمى المتغير الذي يراد التنبؤ بقيمته المتغير التابع ويرمز له بالرمز ص . وتجدر الإشارة الى وجود أسماء أخرى لهذه المتغيرات ، فمثلاً قد يسمى س مُدخلاً ويسمى ص مُخرجاً ، أو قد يسمى س مؤثراً ويسمى ص استجابة ، . . . وهكذا . ويجب على الباحث ، في جميع الأحوال ، أن يكون على دراية كاملة بدور كل من هذه المتغيرات في عملية التنبؤ .

ويعتمد مفهوم التنبؤ الاحصائي على مفهوم الارتباط . يكون المتغيران س ، ص مرتبطان اذا كانت مشاهدات أحدهما تساعد في التنبؤ بمشاهدات الآخر . فمثلاً يقال أن عمر السيارة المستعملة وسعرها مرتبطين اذا أمكن الاعتماد على عمر السيارة للتنبؤ بسعرها . وتجدر الإشارة إلى أن مقاييس الارتباط بين متغيرين تستخدم عادة لدراسة قوة العلاقة بينهما دون الاهتمام بتصنيفهما إلى متغير مستقل وآخر تابع . فمثلاً ، يعتمد على هذه المقاييس

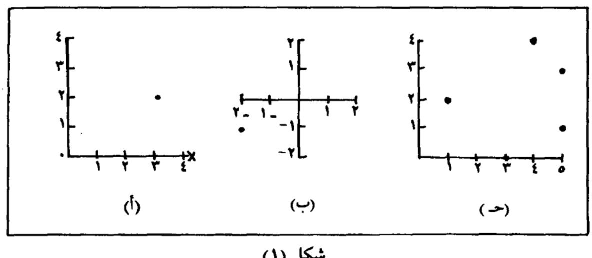
لتحديد المتغيرات الاحصائية التي يمكن أن تغني عن متغيرات أخرى . فإذا كان هناك ارتباط قوي بين متغيرين فإنه يمكن تبسيط الأمور والاستغناء عن أحدهما .

تستخدم مقاييس الارتباط أيضاً لاكتشاف علاقات السببية المحتملة بين المتغيرات المختلفة . ذلك أن ملاحظة ارتباط قوي بين متغيرين يعتبر الخطوة الأولى لدراسة ما إذا كان حدوث أحدهما يمكن أن يعتبر سبباً لحدوث الآخر . فمثلاً ، لوحظ في بلد ما وجود ارتباط قوي بين درجة انتشار التهابات العيون وعدد المسابح العامة في مناطق البلد المختلفة . وقد أدت هذه الملاحظة إلى الاهتمام بدراسة استخدام المسابح العامة كسبب ممكن للإصابة بالتهابات العيون .

وتجدر الإشارة إلى أن وجود ارتباط قوي بين متغيرين لا يعني بالضرورة وجود علاقة سببية بينهما ، اذ لا بد بالاضافة الى ذلك من وجود أساس منطقي يسمح باعتار أحدهما سبباً للآخر . فمثلاً لوحظ وجود ارتباط قوي بين عدد المجانين وعدد أجهزة التليفزيون المستخدمة في بلد ما من عام لآخر . ولا يعني ذلك بالطبع أن هناك علاقة سببية بين هذين المتغيرين حيث لا يوجد أساس منطقي يسمح بهذا الاستنتاج . كذلك يمكن الاعتماد في دراسة السببية على نتائج تجارب مصممة خصيصاً لهذا الغرض ، وذلك في الحالات التي يمكن أن تخضع فيها العلاقة بين المتغيرين للدراسة التجريبية .

٢ ـ مقدمة رياضية

تتعلق هذه المقدمة الرياضية بكيفية رسم النقط والخطوط بيانياً . يبدأ الرسم البياني ، كما هو معروف ، بمحورين ؛ الأفقي ويسمى محور س والرأسي ويسمى محور ص . وتمثل كل نقطة في الشكل قيم س وص المناظرة لها . فمثلاً ، تمثل النقطة التي تظهر في شكل ١ (أ) القيم س = ٣ ، ص = ٢ ، حيث يلاحظ أن النقطة قد وضعت في الشكل عند تقاطع هاتين



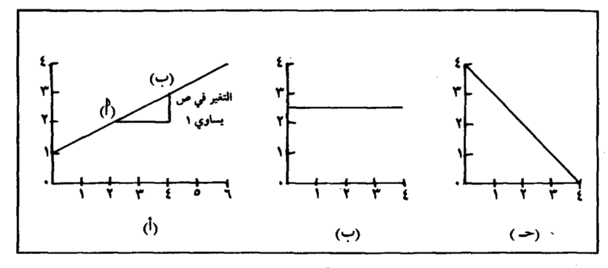
شكل (١)

القيمتين . كذلك تمثل النقطة التي تظهر في شكل ١ (ب) القيم س = ـ ٢ ، ص = ـ ١ . ويمكن أن تكتب مثل هذه القيمة على الشكل (ـ ٢ ، ـ ١) . ويمكن للقارىء على سبيل التمرين تحديد القيم التي تمثلها النقط المختلفة التي تظهر في شكل ١ (جـ) .

يعطي شكل ٢ (أ) خطاً مستقيماً . اذ أخذت أي نقطتين على هذا الخط ولتكن النقط أ ، ب ، فإنه يلاحظ أن هناك تغيراً في قيم س وتغيراً في قيم ص بين هاتين النقطتين . يلاحظ أيضاً أن التغير في قيمة س يساوي ٢ وأن التغير في قيمة ص يساوي١ ، أي أن التغير في قيمة ص يساوي نصف التغير في قيمة س . اذا أخذت أي نقطتين أخريتين على الخط فإن التغير في قيمة ص يساوي دائماً نصف التغير في قيمة س . وتسمى هذه النسبة ميل الخط ، أي أن :

$$\text{ميل الخط} = \frac{\text{التغير في قيمة ص}}{\text{التغير في قيمة س}}$$ ، لأي نقطتين على الخط المستقيم .

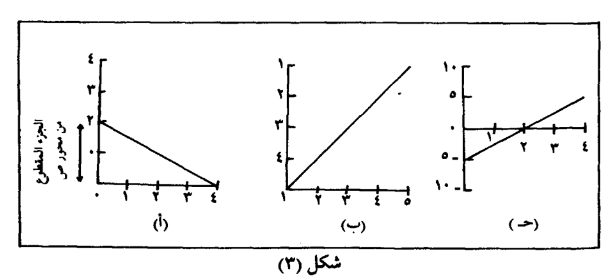
ويمثل ميل الخط المستقيم مقدار التغير الذي يحدث في قيمة ص اذا زادت قيمة س بمقدار وحدة واحدة . فإذا كانت قيمة الميل $= \frac{1}{2}$ فإن ذلك يعني أن التغير في قيمة س بمقدار ١ يتبعه تغير في قيمة ص بمقدار $\frac{1}{2}$. ويلاحظ تبعاً لذلك أن قيمة ميل الخط في شكل ٢ (ب) تساوي الصفر لأن زيادة قيمة س لا



شكل (٢)

يتبعها أي تغير في قيمة ص . كذلك فإن قيمة ميل الخط في شكل ٢ (جـ) تساوي ـ ١ لأن زيادة قيمة س بوحدة واحدة يتبعهـا نقصان قيمـة ص بوحـدة واحدة . اذا كانت قيمة الميل موجبة فإن الخط يكون صاعداً كما في شكل ٢ (أ) . اذا كانت قيمة الميل = صفر فإن الخط المستقيم يكون موازياً للمحـور الأفقي كمـا في شكل ٢ (ب) . أمـا إذا كانت قيمـة الميل سـالبة فـإن الخط المستقيم يكون هابطاً كما في شكل ٢ (جـ) .

ويـلاحظ أن الخط المستقيم يقطع المنحنى الـرأسي عنـد نقـطة مـا ، وتكون قيمة الجزء المقطوع من محور ص مساوية لقيمة ص المنـاظرة لقيمـة س = صفر على هذا الخط . فمثلاً يلاحظ أن الجزء المقطوع من محور ص في الشكل ٣ (أ) يساوي ٢ وأن الجزء المقطوع من محـور ص في الشكل ٣



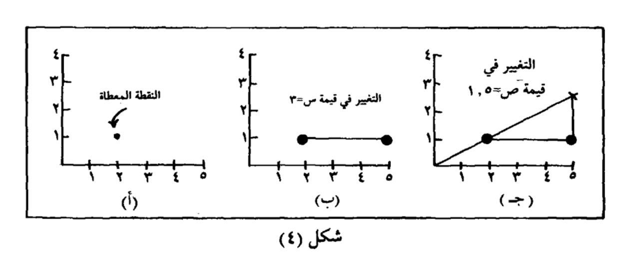
شكل (٣)

(ب) يساوي صفراً وأن الجزء من محور ص في الشكل ٣ (جـ) يساوي − ٥ .

وتوضح الأمثلة التالية كيف رسم الخط المستقيم بيانياً .

مثال (١) : ارسم الخط المستقيم الذي يمر بالنقطة (٢ ، ١) والذي ميله يساوي $\frac{1}{2}$.

الحل : نبدأ برسم المحورين ، ونرسم النقطة المعطاة (٢ ، ١) كما في شكل ٤ (أ) . تحرك مسافة أفقية ملائمة الى يمين هذه النقطة وليكن مقدارها ٣ وضع علامة مناسبة عند نهاية المسافة ، كما في شكل ٤ (ب) . تمثل هذه المسافة مقدار الزيادة في قيمة س. حتى يكون ميل الخط مساوياً $\frac{1}{2}$ ، ارسم عمودا صاعدا من عند العلامة طوله يساوي $1\frac{1}{2}$ ، وضع نقطة في نهايته . اذا وصل بين هذه النقطة والنقطة الأصلية (٢ ، ١) فإن ذلك يؤدي إلى الحصول على الخط المطلوب ، كما في شكل ٤ (جـ) .



شكل (٤)

مثال (٢) اذا كان معلوماً أن ص = $\frac{1}{2}$س + ١ ، فاحسب قيمة ص عندما تكون س = ١ ، ٢ ، ٣ ، ٤ . ارسم النقط التي تحصل عليها . هل تقع جميع هذه النقط على خط مستقيم ؟ إذا كانت الاجابة بنعم فأوجد قيمة ميل الخط وقيمة الجزء المقطوع من محور ص .

الحل :

عندما تكون قيمة س = ١ فإن قيمة ص = ($\frac{1}{2}$ × ١) + ١ = $1\frac{1}{2}$

عندما تكون قيمة س = ٢ فإن قيمة ص = ($\frac{1}{2}$× ٢) + ١ = ٢

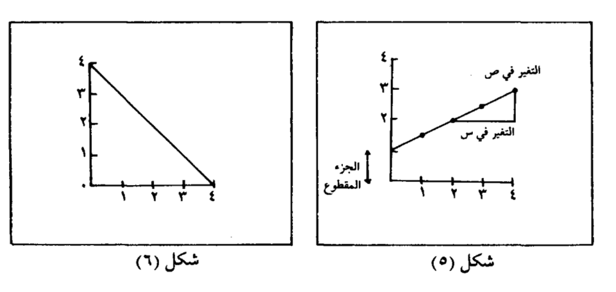
عندما تكون قيمة س = ٣ فإن قيمة ص = ($\frac{1}{2}$× ٣) + ١ = ٢$\frac{1}{2}$

عندما تكون قيمة س = ٤ فإن قيمة ص = ($\frac{1}{2}$× ٤) + ١ = ٣

أي أن النقط هي: (١، ١$\frac{1}{2}$)، (٢، ٢)، (٣، ٢$\frac{1}{2}$)؛ (٤، ٣).

وتظهر هذه النقط في شكل (٥) حيث يلاحظ وقوعها جميعاً على خط مستقيم. ويلاحظ أن ميل الخط =$\frac{1}{2}$وأن الجزء المقطوع من محور ص يساوي ١ . ويلاحظ كذلك أن قيمة الميل تساوي قيمة معامل س في المعادلة وأن قيمة الجزء المقطوع من محور ص يساوي قيمة الحد الثابت في المعادلة.

ويعتبر هذا المثال حالة خاصة لقاعدة عامة تتمثل في أن الرسم البياني لأي معادلة على الشكل ص = أ س + ب يأخذ شكل خط مستقيم ميله = أ والجزء الذي يقطعه من محور ص يساوي ب .



شكل (٦) شكل (٥)

مثال (٣) ما هي معادلة الخط المستقيم الذي يظهر في شكل (٦) ؟

الحل: يلاحظ أن ميل هذا الخط = -١ وأن الجزء المقطوع من محور ص يساوي ٤ ، وعلى ذلك فإن المعادلة هي ص = ـ س+٤ .

مثال (٤): ارسم الخط المستقيم الذي معادلته هي ص = $-\frac{١}{٢}$س + ٤ .

الحل : هذا خط مستقيم ميله = $-\frac{١}{٢}$ والجزء الذي يقطعه من محور ص = ٤ . ولما كان من المعروف أن رسم الخط المستقيم يتحدد بأي نقطتين عليه فإنه يمكن استخدام المعادلة المعطاة لتحديد نقطتين على الخط . فمثلاً عند س = صفر نجد أن ص = ٤ .

وعند س = ٢ نجد أن ص = ٣ .

وعلى ذلك ترسم النقطتان (صفر ، ٤) ، (٢ ، ٣) وينتج الخط المطلب بالتوصيل بينهما ، كما يظهر في شكل (٧) .

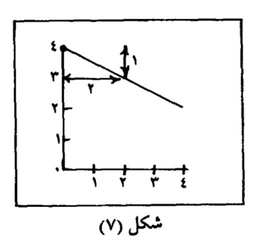
ويمكن للقارىء على سبيل التمرين أن يرسم الخطوط التالية :

(أ) ص = ٢ س + ١ .

(ب) ص = $\frac{١}{٢}$س + ٢

(جـ) ص = $-\frac{١}{٤}$س + ٤ .

(ء) ص = $\frac{٣}{٤}$س − ١ .



شكل (٧)

٣ ـ شكل الانتشار

تتألف مجموعة البيانات المزدوجة من مشاهدات عن متغيرين س ، ص . وتكتب المشاهدة الخاصة بكل مفردة من مفردات الدراسة بأسلوب يوضح ذلك مثل (٧ ، ٢) أو (٤٨ ، ١٥) حيث يشير الرقم الأول إلى قيمة س للمفردة ويشير الرقم الثاني إلى قيمة ص لنفس المفردة . اذا كان عدد المشاهدات كبيراً ، فإنه يكون من الصعب التعرف على خصائص توزيع كل من المتغيرين أو نمط العلاقة بينهما بالاعتماد على هذه البيانات الخام . وقد سبقت الاشارة إلى أهمية تنظيم وعرض البيانات الخام كخطوة أولى من خطوات التحليل الاحصائي .

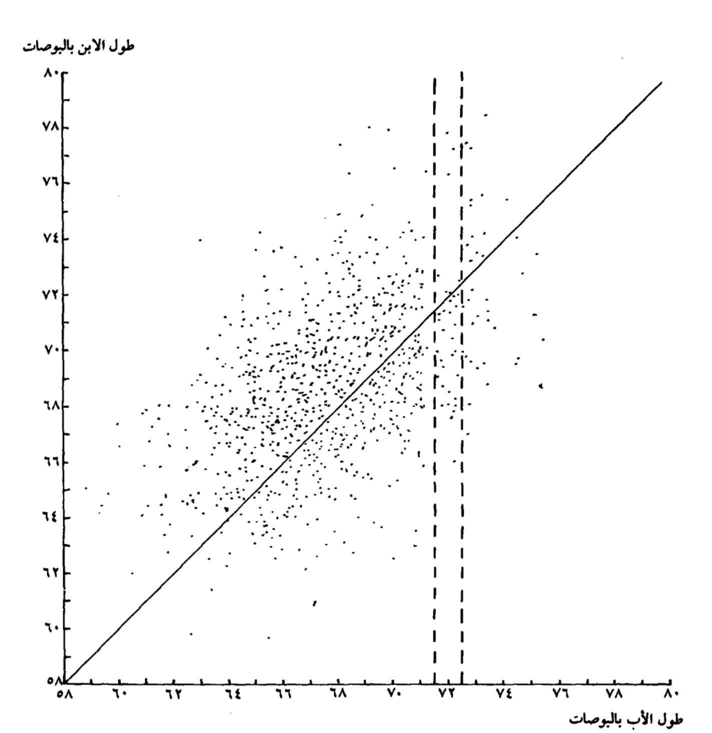
يعتبر شكل الانتشار أحد الأساليب الهامة لعرض البيانات المـزدوجة . ويرسم هذا الشكل بتمثيل المشاهدة الخاصة بكل مفردة بنقطة توضع عند تقاطع قيمتي س ، ص للمفـردة . وتوضـح الأمثلة التاليـة الاستخدامـات المختلفة لشكل الانتشار .

مثـال (٥) : ترجـع الأصول التـاريخية لـلأساليب المختلفـة لـدراسـة العلاقات بين المتغيرات الاحصائية الى أبحاث الوراثة في القرن التاسع عشر التي كـانت تهتم بوصف العـلاقة بين خصـائص الآباء وخصـائص أبنـائهم . جمعت في أحد هذه الأبحاث بيانات من ١٠٧٨ أسرة عند طول الأب وطـول أحد أبنائه (مقاساً بالبوصات) . ويبدأ التحليل الاحصائي لهذا العـدد الكبير من المشاهدات بتنظيمها وعرضها في شكل بياني مناسب .

يعطي شكل (٨) شكل الانتشار المناظر لهذه البيانـات . ويتطلب رسم هـذا الشكل تحـديد المتغيـر المستقل والمتغيـر التابـع في المتغيـرين محـل الدراسة . ولما كان طول الأب يؤثر في تحديد طول الابن وليس العكس فقد اعتبر طول الأب كمتغير مستقل وأخـذ على المحور الأفقي س ، بينمـا اعتبر طول الابن متغيراً تابعاً يظهر على المحور الرأسي ص . وتمثل كل نقـطة في شكل الانتشار قيمـة س وقيمة ص لأسـرة واحدة ، أي أن الشكـل يتألف من ١٠٧٨ نقطة .

يلاحظ من شكل الانتشار وجود اتجاه عام في البيانات نحو الصعود بميل موجب . أي أن قيمة المتغير ص تتجه نحو الارتفاع كلما زادت قيمة المتغير س . ويقـال في هذه الحـالة أن هنـاك ارتباطـاً موجبـاً بين طول الأب وطـول الابن ، أي أن الأب طويل القامة يكون إبنه في الغالب طويل القامة أيضاً .

لا يقتصر شكل الانتشار على توضيح اتجاه الارتباط بين المتغيرين س ، ص ، بل يتعداه الى وصف درجة قوة هذا الارتباط . فمثلاً ، اذا كان الارتباط بين طول الأب وطول الابن قوياً فإنه يتوقع أن يكون طول الابن قريب من طول أبيـه ، أي أن تكون قيمـة ص قريبـة من قيمة س . وبعبـارة أخرى ، تكـون

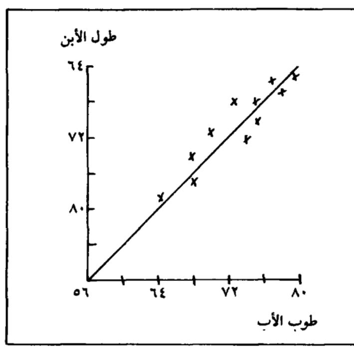


شكل (٨) : شكل الانتشار للعلاقة بين طول الأب وطول ابنه
(المصدر : Freeman et al. انظر قائمة المراجع المختارة)

العلاقة بين طول الابن وطول أبيه قريبة من خط مستقيم معادلته هي ص = س ، وتكون نقط شكل الانتشار مركزة بالقرب من هذا الخط كما يظهر في شكل (٩) .

يلاحظ وجود اختلافات واضحة بين شكل (٩) وشكل (٨) ، اذ توجد مسافات كبيرة بين النقط والخط المستقيم ص = س في شكل الانتشار الفعلي للبيانات . ويدل ذلك على وجود فروق كبيرة نسبياً بين أطوال الابناء وأطوال

آبائهم ، مما يعني أن الارتباط بين طول الأب وطول الابن ضعيفة . ويترتب على ذلك أن استخدام طول الأب كأساس للتنبؤ بطول ابنه يكون عرضة لأخطاء واضحة . فمثلاً ، اذا أريد التنبؤ بطول ابن اذا علم أن طول أبيه يساوي ٧٢ بوصة ، فإنه يلاحظ من شكل الانتشار ارتفاع درجة التشتت في أطوال هؤلاء الأبناء . ولا يجب في هذه الحالة الاكتفاء بقيمة وحيدة للتنبؤ لأن مثل هذه القيمة تكون قليلة الكفاءة كمقياس للموضع في ضوء التشتت الواسع للبيانات .



شكل (٩) : مثال لعلاقة قوية بين طول الأب وطول ابنه

ونخلص من هذا المثال إلى ما يلي :

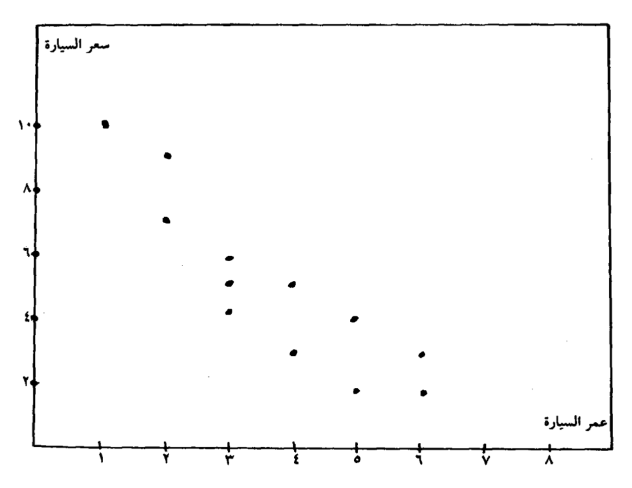
(أ) أن شكل الانتشار يوضح طبيعة العلاقة بين المتغيرين س ، ص . ويشمل ذلك وصف اتجاه وقوة الارتباط بينهما .

(ب) يمكن الاعتماد على قيم المتغير س للتنبؤ بقيم المتغير ص ، إذا كان الارتباط بينهما قوياً . أما إذا كان هذا الارتباط ضعيفاً فإن عملية التنبؤ تكون منخفضة الكفاءة وعرضة لأخطاء كبيرة .

مثال (٦) : جمعت البيانات التالية عن عمر السيارة بالسنوات وسعر بيع السيارة بآلاف الدراهم لعينة من ١٢ سيارة بيعت مؤخراً في سوق الحراج .

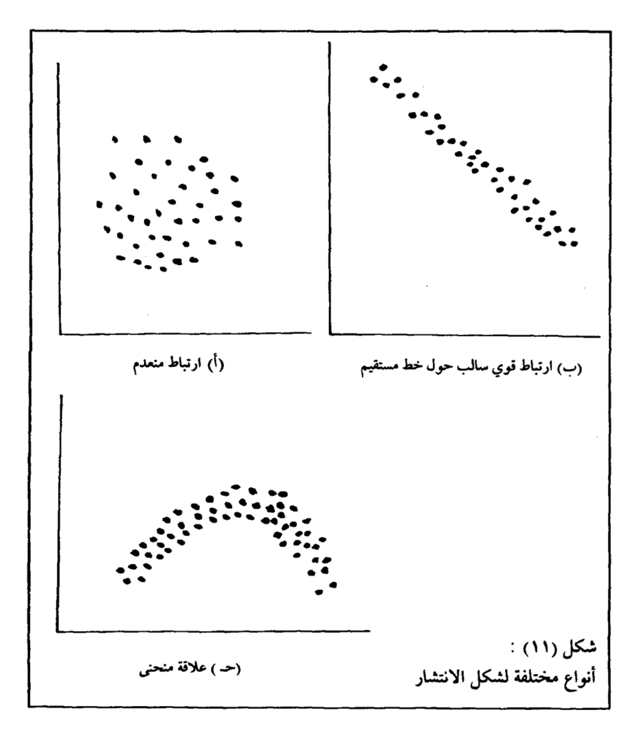
رقم السيارة :	١	٢	٣	٤	٥	٦	٧	٨	٩	١٠	١١	١٢
عمر السيارة (س) :	٣	٥	٢	٦	٤	٦	٣	١	٢	٣	٤	٥
سعر السيارة (ص) :	٦	٤	٧	٣	٣	٢	٥	١٠	٩	٤	٥	٢

يظهر شكل الانتشار المناظر لهذه البيانـات في شكل (١٠) . ويـلاحظ اتجاه البيانات في الشكل نحو الهبوط بميل سالب ، أي أن قيمـة المتغير ص تقل كلما زادت قيمة المتغير س . ويقال في هذه الحالة أن هناك ارتباطاً سالباً بين عمر السيارة وسعر بيعها . ويتفق ذلك مع مـا هو متـوقع من أن سعـر بيع السيارة المستعملة يقل مـع قدم السيارة . ويـلاحظ كذلـك وجود انحنـاء في البيانات ، اذ يهبط سعر السيارة بسرعة خلال السنوات الأولى من العمر ، ثم يقل معدل هذا الهبوط بعد ذلك اذ ربما يكون سعر بيع السيارة أقل حسـاسية لعمرها بعد حد معين .



شكل (١٠) : شكل الانتشار للعلاقة بين عمر السيارة وسعرها

ويلاحظ عند رسم شكل الانتشار أن نقطة تقاطع محور س مع محور ص قد تكون النقطة (٠ ، ٠) كما في شكل (١٠) ، أو قد تـكون نقطة أخرى كما في شكل (٨) . ويتحدد مكان هذه النقطة تبعاً لقيم المشاهدات المعطاة . اذ لا يجب البدء عند الصفر اذا كانت قيم جميع المشاهـدات بعيدة عنـه وذلك



(أ) ارتباط منعدم

(ب) ارتباط قوي سالب حول خط مستقيم

(حـ) علاقة منحنى

شكل (١١) :
أنواع مختلفة لشكل الانتشار

تفادياً لإهدار وعدم استخدام المساحة المستخدمة لرسم الشكل .

يعطي شكل (١١) أمثلة لأنواع أخرى لشكل الانتشار . اذ يوضح شكل (١١ ـ أ) الحالة التي يكون فيها الارتباط منعدماً بين س ، ص ، بينما يمثل شكل (١١ ـ ب) حالة وجود ارتباط سالب قوي حول خط مستقيم بين س ، ص ، ويوضح شكل (١١ ـ جـ) مثالاً لعلاقة منحنى بين س ، ص .

ويجب الإشارة إلى ضرورة مراعاة القواعد العامة للرسم البياني عند

إعداد شكل الانتشار ، وبصفة خاصة ينبغي مراعاة الآتي :

أ ـ ضرورة وضع عنوان واضح وموجز للشكل .

ب ـ ضرورة ظهور اسم المتغير الذي يمثل على كل محور مع ذكر وحدات القياس .

جـ ـ اعطاء مصدر البيانات ، كلما كان ذلك ملائماً .

د ـ يظهر المتغير المستقل س على المحور الأفقي بينما يمثل التابع ص على المحور الرأسي .

هـ ـ يجب اختيار مقياس رسم مناسب يسمح بأن يظهر شكل الانتشار في معظم المساحة المخصصة للرسم . وفي هذا الصدد يمكن بدء المقياس على أي من المحورين عند نقطة أخرى غير الصفر .

٤ ـ معامل الارتباط الخطي

نتتقل الآن إلى مناقشة كيفية استخدام مقاييس عددية لوصف السمات العامة لشكل الانتشار . يلاحظ أن شكل الانتشار يأخذ شكل سحابة من النقط . ويمكن وصف السحابة عددياً بأسلوب ملائم باتباع الخطوات التالية :

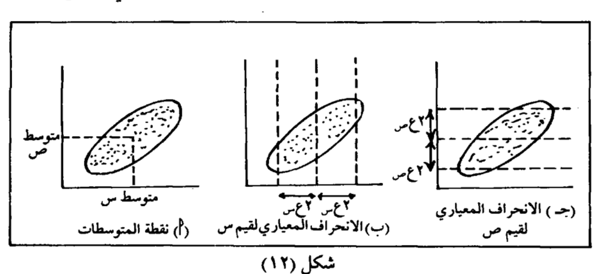
أ ـ تحديد مركز السحابة . ويتم ذلك بحساب متوسط المتغير س (أي $\overline{س}$) ومتوسط قيم المتغير ص (أي $\overline{ص}$) . وتكون نقطة المتوسطات ($\overline{س}$ ، $\overline{ص}$) هي مركز السحابة . انظر شكل (١٢ ـ أ) .

ب ـ قياس درجة التشتت الأفقي في السحابة . ويتم ذلك بحساب قيمة الانحراف المعياري لقيم المتغير س (أي $ع_{س}$) . ويتوقع طبقاً للقاعدة العملية أن تقع معظم نقط السحابة داخل مسافة تقل عن انحرافين معياريين من الوسط الحسابي . انظر شكل (١٢ ـ ب) .

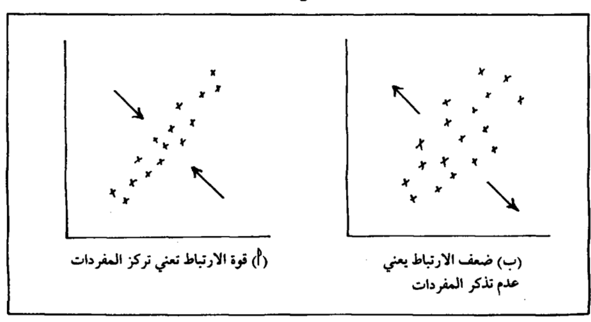
جـ ـ قياس درجة التشتت الرأسي في السحابة . ويتم ذلك بحساب قيمة الانحراف المعياري لقيم المتغير ص (أي $ع_{ص}$) . ويتوقع أن تقع معظم

نقط السحـابة داخـل مسـافـة تقـل عن انحـرافين معيـاريين من الـوسط الحسابي . انظر شكل (١٢ ـ جـ) .

د ـ حساب مقياس اضافي يقيس قوة العلاقة بين المتغيرين داخل السحـابة . وتتضح أهمية هذا المقياس بالنظر إلى شكـل (١٣) . اذ يلاحظ وجـود سحابتين لهما نفس المركز ونفس درجة التشتت الأفقي والرأسي ، ومع ذلك يختلفان في شكلهما العام . تتسم نقط السحابة في شكل (١٣ ـ أ) بتركزها الواضح حول خط مستقيم ، مما يدل على وجود علاقـة خطيـة قوية بين المتغيرين ، على حين يبدو أن العلاقة أقل قوة في شكل (١٣ ـ



شكل (١٢)



شكل (١٣)

ب) نتيجة توزيع النقط حول خط مستقيم على نحو أكثر اتساعاً. ويعني ذلك أن السحابتين يختلفان في قوة العلاقة بين المتغيرين . ويحسب معامل الارتباط لقياس قوة هذه العلاقة .

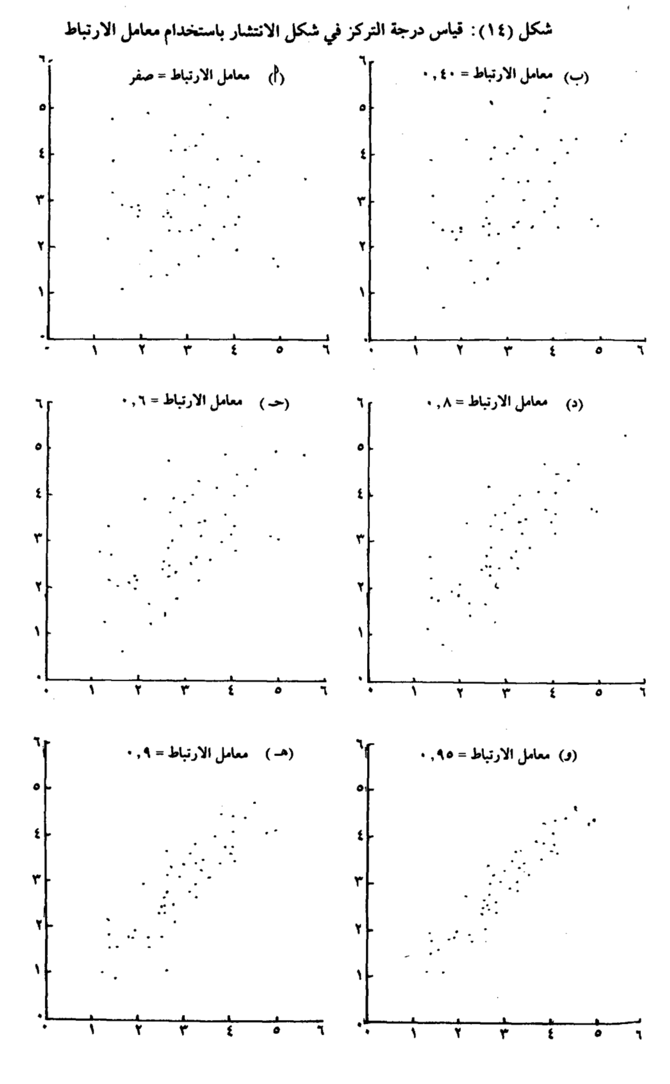
وبعبارة أخرى فإن معامل الارتباط هو مقياس لقوة العلاقة الخطية بين متغيرين . ويمكن تبعاً لذلك تلخيص السمات الأساسية لمجموعات بيانات مزدوجة باستخدام خمس مقاييس هي الوسط الحسابي لقيم س والوسط الحسابي لقيم ص والانحراف المعياري لقيم س والانحراف المعياري لقيم ص ومعامل الارتباط بين قيم س وقيم ص .

ويقيس معامل الارتباط اتجاه وقوة العلاقة الخطية بين متغيرين . وتقع قيمته داخل المدى (- ١ ، + ١) . اذا كانت قيمة المعامل موجبة (أي تقع بين صفر ، ١) فإن هذا الارتباط الموجب يعني أن سحابة النقط في شكل الانتشار تكون صاعدة بميل موجب . ويدل ذلك على أن زيادة قيم أحد المتغيرين يتبعها زيادة في قيم المتغير الآخر . أما إذا كانت قيمة المعامل سالبة (أي تقع بين - ١ ، صفر) فإن سحابة النقط في شكل الانتشار تكون هابطة بميل سالب . ويدل ذلك على أن زيادة قيم أحد المتغيرين يصاحبها نقصان في قيمة المتغير الآخر . ويمكن تفسير معنى القيمة العددية لمعامل الارتباط بالاستعانة بأشكال الانتشار التي تظهر في شكلي (١٤) ، (١٥) .

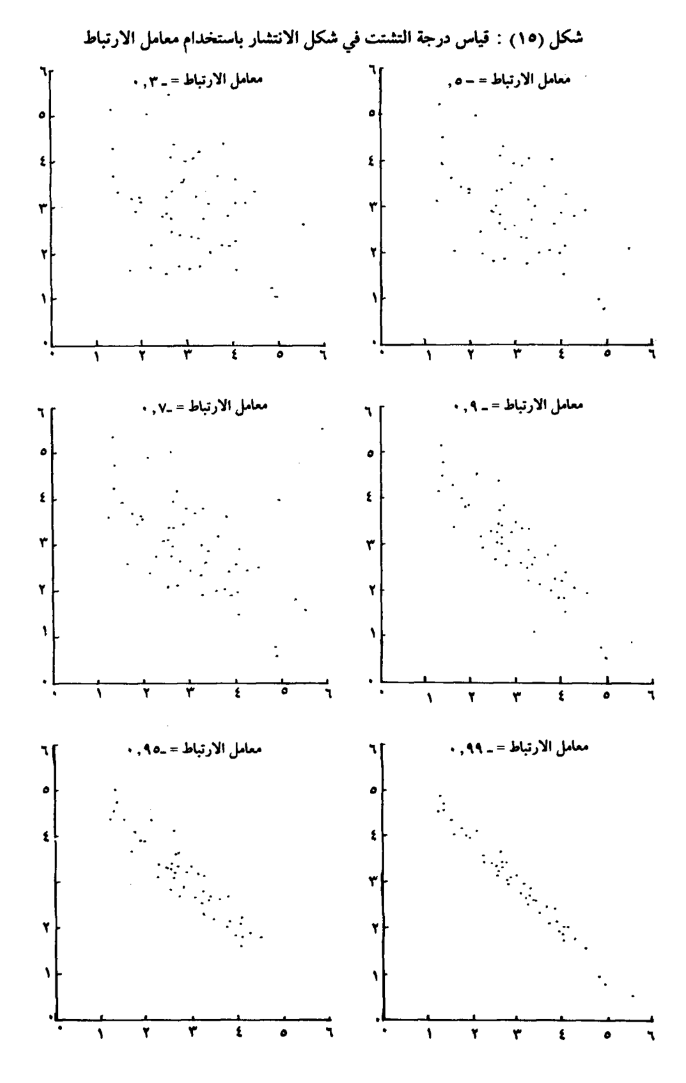
يعطي شكل (١٤) ستة أشكال انتشار مختلفة تتفق جميعها في قيم المتوسطات والانحرافات المعيارية للمتغيرين س، ص وتختلف فقط في قوة العلاقة بين المتغيرين . وقد تم إنشاء هذه الأشكال على الحاسب الآلي باستخدام برنامج حسابي مناسب .

يمثل شكل (١٤ ـ أ) شكل الانتشار عندما تكون قيمة معامل الارتباط مساوية الصفر ، حيث يلاحظ أن زيادة قيم س لا يصاحبها تغيير في نمط قيم ص . وينعدم الارتباط بين س ، ص في هذه الحالة إذ لا يوجد أي اتجاه في النقط نحو التركز حول خط مستقيم . ويناظر شكل (١٤ ـ ب) قيمة معامل

شكل (١٤) : قياس درجة التركز في شكل الانتشار باستخدام معامل الارتباط



شكل (١٥) : قياس درجة التشتت في شكل الانتشار باستخدام معامل الارتباط



ارتبـاط تساوي ٤ , ٠ ويتسم هـذا الشكل ببـدء ظهور اتجـاه خطي طفيف في البيانات ، ولكن الارتباط بين المتغيرين لا يـزال ضعيفاً . كـذلك فـإن شكل (١٤ ـ جـ) يمثـل قيمة معـامل ارتبـاط تسـاوي ٦ , ٠ حيث يـلاحظ أن نمط العلاقة الخطية أصبح أكثر وضوحاً في البيانات . وهكذا بالنسبة لباقي الأشكال حيث يلاحظ أنه كلما اقتربت قيمة المعامل من الواحد الصحيح كلما دل ذلك على قوة العلاقة الخطية بين المتغيرين وارتفـاع درجة تـركز النقط حـول خط مستقيم . إذا كانت قيمة معامل الارتبـاط تسـاوي الواحـد ، دل ذلـك على وجـود ارتبـاط تـام بين المتغيرين . وتقـع جميـع نقط شكل الانتشـار في هـذه الحالة على خـط مستقيم .

يعطي شكل (١٥) أمثلة لعلاقات ارتباط سالب تختلف في درجة قوتها . وتظهر الإشارة (ـ) في معامـل الارتباط للدلالـة على وجود ارتبـاط سالب . ويتضح من هذه الأشكال أنه كلما اقتربت قيمة معامل الارتباط من ـ ١ دل ذلك على قوة العلاقة الخطية بين المتغيرين وعلى ارتفاع درجة تركز النقط حول خط مستقيم . اذا كانت قيمة معامل الارتباط تساوي ـ ١ فإن ذلك يعني وجود علاقة خطية تامة بين المتغيرين . وتقع جميع نقط شكل الانتشار في هذه الحالة على خط مستقيم يكون ميله سالباً .

وتشيـر هذه الـرسوم البيـانية إلى قـاعدة تقـريبية يعتمـد عليها كثيـر من الباحثين لوصف قوة العلاقـة بين متغيرين على أسـاس قيمة معـامل الارتبـاط بينهما . إذا كانت القيمة المطلقة لمعامل الارتباط تزيد عن ٨ , ٠ فإن العلاقة بين المتغيرين تكون قوية . وتكون العلاقة متوسطة اذا تراوحت قيمة المعامل بين ٥ , ٠ ، ٨ , ٠ وتكـون العلاقـة ضعيفة إذا كـانت القيمة المـطلقة لمعـامل الارتباط أقل من ٥ , ٠ ويتضح ذلك في شكل (١٦) .



شكل (١٦) : تفسير قيمة معامل الارتباط

٥ - خط الارتباط التام

يزداد تركز نقط شكل الانتشار حول خط مستقيم كلما اقتربت قيمة معامل الارتباط من الواحد . وتقع جميع النقط على هذا الخط تماماً في حالة العلاقة التامة التي تكون قيمة معامل الارتباط فيها مساوية للواحد . ويعتبر خط الارتباط التام تبعاً لذلك معياراً يمكن استخدامه في تفسير قوة العلاقة بين متغيرين ، بمعنى أن هذه العلاقة تكون اكثر قوة كلما تركزت نقط شكل الانتشار حول هذا الخط . ويتطلب ذلك بالطبع تحديد هذا الخط .

يمكن شرح مفهوم هذا الخط باستخدام مثال عن العلاقة بين طول الشخص ووزنه في أحد المجتمعات . عندما يقال أن هناك ارتباطاً قوياً بين طول الشخص ووزنه فإن ذلك يعني أنه اذا كان هناك شخص طوله قريب من متوسط الأطوال فإنه وزنه أيضاً يتوقع أن يكون قريباً من متوسط الأوزان ، أما إذا كان طوله يزيد عن متوسط الأطوال فإن وزنه أيضاً يزيد عن متوسط الأوزان . وهكذا . وبعبارة أخرى فإن الارتباط القوي بين طول الشخص ووزنه يعني وجود اتجاه قوي للتناظر بين الوضع النسبي لوزن الشخص بين الاوزان والوضع النسبي لطوله بين الأطوال . ويكون الارتباط تاماً إذا كان هذا التناظر تاماً .

سبقت الإشارة إلى أن المكانة النسبية لمشاهدة ما ضمن مجموعة من المشاهدات تتحدد بتحويل المشاهدات إلى وحدات معيارية ، أي بقياس اختلاف كل مشاهدة عن الوسط الحسابي بوحدات من الانحرافات المعيارية . وعلى ذلك فإن الارتباط التام بين متغيرين س ، ص يحدث اذا كانت كل درجة معيارية للمتغير س يناظرها درجة معيارية مساوية للمتغير ص . أي أن العلاقة بين س ، ص في هذه الحالة يمكن أن تكتب على الشكل :

$$\frac{ص - \overline{ص}}{ع_{ص}} = \frac{س - \overline{س}}{ع_{س}}$$

أو $ص - \bar{ص} = \frac{ع_ص}{ع_س} (س - \bar{س})$

أو $ص = \frac{ع_ص}{ع_س} \times س + \left(\bar{ص} - \frac{ع_ص}{ع_س} \times \bar{س} \right)$

وهذه معادلـة خط مستقيم يمر بنقـطة المتوسـطات ($\bar{س}$ ، $\bar{ص}$) وميله يسـاوي $\frac{ع_ص}{ع_س}$ ويسمى هـذا الخط بخط الارتبـاط التـام أو خط الانحــرافـات المعيارية .

تجدر الإشارة إلى أن ميل هذا الخط يكون سالباً إذا كانت العلاقة التامة عكسية ، ويكون الميل في هذه الحالة مساوياً $- \frac{ع_ص}{ع_س}$

٦ ـ طرق حساب معامل الارتباط

سوف نرمز لمعامل الارتباط بين متغيرين بالرمز ر . تحسب قيمة معامل الارتباط ر باتباع الخطوات البسيطة التالية :

(أ) تحول مشاهدات المتغير س إلى وحدات معياريـة ، ولنرمـز لها بـالرمـز $ى_س$.

(ب) تحول مشاهدات المتغير ص إلى وحدات معيارية ، ولنرمـز لها بـالرمـز $ى_ص$.

(حـ) تضرب كل وحدة معيارية للمتغير س بالوحدة المعيارية المناظرة للمتغير ص .

(د) تجمع حواصل الضرب ويقسم الناتج على (ن ـ ١) ، حيث تمثـل ن عدد المشاهدات ، لتنتج قيمة معامل الارتباط ر .

أي أن $ر = \frac{1}{ن - 1} \sum ى_س . ى_ص$.

مثال (٧) : احسب معامل الارتباط بين المتغيرين س ، ص باستخدام البيانات التالية :

س :	١	٣	٤	٥	٧
ص :	٥	٩	٧	١	١٣

الحل :

نحسب قيم $\bar{س}$ ، $\bar{ص}$ ، $ع_{س}$ ، $ع_{ص}$ حيث يلاحظ أن :

$$\bar{س} = \frac{١+٣+٤+٥+٧}{٥} = \frac{٢٠}{٥} = ٤$$

$$ع^{٢}_{س} = \frac{١}{ن-١} مجـ (س - \bar{س})^{٢} = \frac{١}{٥-١}$$

$$[(١-٤)^{٢} + (٣-٤)^{٢} + (٤-٤)^{٢} + (٥-٤)^{٢} + (٧-٤)^{٢}]$$

$$= \frac{١}{٤} [٩+١ + ٠+١ + ٩] = ٥$$

وبالتالي تكون قيمة $ع_{س} = \sqrt{٥} = ٢,٢٣٦$

كذلك فإن $\bar{ص} = \frac{٥+٩+٧+١+١٣}{٥} = \frac{٣٥}{٥} = ٧$

$$ع^{٢}_{ص} = \frac{١}{٤} [٨٠] = ٢٠$$

وبالتالي تكون قيمة $ع_{ص} = \sqrt{٢٠} = ٤,٤٧٢$

ثم تستخدم هذه المقاييس لحساب الدرجات المعيارية ومعامل الارتباط كما في الجدول التالي :

س	ص	الوحدات المعيارية لقيم س ى س	الوحدات المعيارية لقيم ص ي س	حاصل الضرب ي س × ي ص
١	٥	(١ - ٤) / ٢,٢٣٦ = ـ ١,٣٤	(٥ - ٧) / ٤,٤٧٢ = ـ ٠,٤٥	ـ ١,٣٤ × ـ ٠,٤٥ = ٠,٦٠
٣	٩	(٣ - ٤) / ٢,٢٣٦ = ـ ٠,٤٥	(٩ - ٧) / ٤,٤٧٢ = ٠,٤٥	ـ ٠,٤٥ × ٠,٤٥ = ـ ٠,٢٠
٤	٧	(٤ - ٤) / ٢,٢٣٦ = صفر	(٧ - ٧) / ٤,٤٧٢ = صفر	صفر × صفر = صفر
٥	١	(٥ - ٤) / ٢,٢٣٦ = ٠,٤٥	(١ - ٧) / ٤,٤٧٢ = ـ ١,٣٤	٠,٤٥ × ـ ١,٣٤ = ـ ٠,٦٠
٧	١٣	(٧ - ٤) / ٢,٢٣٦ = ١,٣٤	(١٣ - ٧) / ٤,٤٧٢ = ١,٣٤	١,٣٤ × ١,٣٤ = + ١,٨٠

مجـ ى$_{س}$ × ى$_{ص}$ = ١,٦٠

وتكون قيمة معامل الارتباط ر هي :

$$ر = \frac{1}{ن - 1} \text{ مجـ ى}_{س} \cdot \text{ى}_{ص}$$

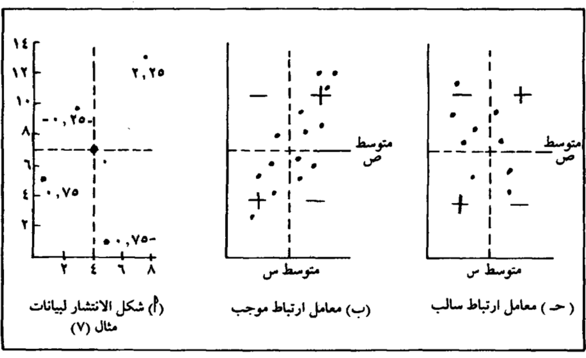
$$= \frac{1}{5 - 1} (1,60) = 0,40$$

ويمكن للقارىء أن يفسر معنى هذه النتيجة برسم شكل الانتشار المناظر للبيانات وملاحظة وجود علاقة خطية ضعيفة بين المتغيرين س ، ص .

يتضح من أسلوب حساب معامل الارتباط أن هذا المعامل خال من وحدات القياس أي لا يوجد له تمييز ، ومن هنا جاءت تسميته بمعامل . ويرجع ذلك إلى اعتماده على الوحدات المعيارية للمتغيرين محل الدراسة ، وهي قيم لا تمييز لها كما هو معروف . ويترتب على ذلك أن تغيير وحدات قياس أي من المتغيرين لا يؤثر في قيمة معامل الارتباط بينهما . ولما كان تغيير وحدات القياس لا يمكن أن يتم إلا من خلال عمليات جمع أو ضرب ، فإن ذلك يعني أن اضافة نفس الرقم لمشاهدات أحد المتغيرين لا يؤثر في قيمة معامل الارتباط . كذلك إذا ضربت جميع المشاهدات في نفس الرقم الموجب فإن ذلك لا يغير من قيمة معامل الارتباط .

ويوضح أسلوب الحساب كيفية قيام معامل الارتباط بقياس قوة العلاقة الخطية بين متغيرين. يعطي شكل ١٧ (أ) شكل الانتشار المناظر لبيانات مثال (٧) ، مع قيمة حاصل ضرب الوحدات المعيارية المناظرة لكل نقطة من نقط هذا الشكل . رسم أيضاً خطان متعامدان يتقاطعان عند نقطة المتوسطات ، ويقسمان شكل الانتشار إلى أربع أجزاء . إذا وقعت نقطة في الجزء الأيسر الأسفل فإن ذلك يعني أن قيم س ، ص المناظرة لهذه النقطة تكون أقل من متوسطاتها وتكون الوحدات المعيارية المناظرة لهما سالبة ، أي أن قيمة حاصل ضرب هذه الوحدات تكون موجبة . كذلك فإن حاصل ضرب الوحدات المعيارية المناظرة لأي نقطة تقع في الجزء الأيمن الأعلى يكون أيضاً موجباً .

بينما تكون قيم حاصل ضرب الوحدات المعيارية المناظرة لأي نقطة تقـع في الجـزء الأيسر ، الأعلى أو الجـزء الأيمن الأسفل سـالبـة . ويحسب معـامـل الارتباط بجمع قيم حواصل الضرب هذه ، سواء كانت سالبة أو موجبة . وعلى ذلك فإن القيمـة الموجبـة لمعامـل الارتباط تـدل على وجود تـركز للنقط في الربعين الموجبين لشكل الانتشار كما يظهر في شكل ١٧ (ب). أما اذا كانت قيمة معامل الارتباط سـالبة فـإن ذلك يعني تـركز النقط في الـربعين السالبين لشكل الانتشار ، كما يتضح في شكل ١٧ (جـ) .



شكل (١٧) : تفسير معامل الارتباط

يمكن تسهيـل العمل الحسـابي اللازم لايجـاد قيمـة معـامـل الارتبـاط بملاحظة أن :

$$\text{ر} = \frac{1}{\text{ن} - 1} \text{ مجـ } \text{ى}_{\text{س}} \cdot \text{ى}_{\text{ص}}$$

$$\text{ر} = \frac{1}{\text{ن} - 1} \text{ مجـ } \left(\frac{\text{س} - \bar{\text{س}}}{\text{ع}_{\text{س}}} \right) \left(\frac{\text{ص} - \bar{\text{ص}}}{\text{ع}_{\text{ص}}} \right)$$

ولما كانت قيم $ع_{س}$ ، $ع_{ص}$ ثابتة ويمكن أخذها خارج علامة المجموع ، فإن ذلك يعني أن :

$$ر = \frac{مجـ (س - \bar{س}) (ص - \bar{ص})}{(ن - ١) ع_{س} ع_{ص}}$$

ويوضح المثال التالي كيفية تطبيق هذه الصيغة الرياضية لحساب ر .

مثال (٨) : استخدم الصيغة الرياضية السابقة لحساب قيمة معامل الارتباط ر بين س ، ص للبيانات المعطاة في مثال (٧) .

الحل : يحتاج تطبيق هذه الصيغة الرياضية إلى الحسابات التي تظهر في الجدول التالي :

س	ص	$س - \bar{س}$	$ص - \bar{ص}$	$(س - \bar{س})^٢$	$(ص - \bar{ص})^٢$	$(س - \bar{س}) \times (ص - \bar{ص})$
١	٥	١-٤ = -٣	٥-٧ = -٢	٩	٤	٦
٣	٩	٣-٤ = -١	٩-٧ = ٢	١	٤	-٢
٤	٧	٤-٤ = ٠	٧-٧ = ٠	٠	٠	٠
٥	١	٥-٤ = ١	١-٧ = -٦	١	٣٦	-٦
٧	١٣	٧-٤ = ٣	١٣-٧ = ٦	٩	٣٦	١٨
٢٠	٣٥	صفر	صفر	٢٠	٨٠	١٦

حيث يلاحظ أن :

$$ع^٢_{س} = \frac{١}{ن - ١} مجـ (س - \bar{س})^٢ = \frac{٢٠}{٤} = ٥$$ وبالتالي فإن $ع_{س} = \sqrt{٥} = ٢,٢٣٦$ ،

$$ع^٢_{ص} = \frac{١}{ن - ١} مجـ (ص - \bar{ص})^٢ = \frac{٨٠}{٤} = ٢٠$$ وبالتالي فإن $ع_{ص} = \sqrt{٢٠} = ٤,٤٧٢$

كذلك فإن مجـ $(س - \bar{س}) (ص - \bar{ص}) = ١٦$. وعلى ذلك فإن قيمة معامل الارتباط ر تكون :

$$ر = \frac{١٦}{٤ \times ٢,٢٣٦ \times ٤,٤٧٢} = \frac{١٦}{٤٠,٠} = ٠,٤$$

ويمكن كذلك حساب قيمة معامل الارتباط ر بأسلوب آخر يعتمد على ملاحظة أن :

$$مجـ (س - \bar{س})^٢ = مجـ س^٢ - \frac{(مجـ س)^٢}{ن} \quad ، \quad وأن$$

$$مجـ (ص - \bar{ص})^٢ = مجـ ص^٢ - \frac{(مجـ ص)^٢}{ن}$$

وهي نتائج سبقت الإشارة إليها عند مناقشة طرق حساب التباين والانحراف المعياري . كذلك يمكن بنفس الطريقة إثبات أن :

$$مجـ (س - \bar{س}) (ص - \bar{ص}) = مجـ س ص - \frac{(مجـ س) (مجـ ص)}{ن}$$

وعلى ذلك فإنه يمكن إيجاد قيمة معامل الارتباط بحساب القيم مجـ س ،مجـ ص ،مجـ $س^٢$ ، مجـ $ص^٢$ ،مجـ س ص كما يتضح من المثال التالي :

مثال (٩) : أعد حل المثال رقم (٨) باستخدام العلاقات المشار إليها أعلاه .

الحل : يمكن تنظيم العمل الحسابي في جدول يشبه الجدول التالي :

س	ص	س ص	$س^٢$	$ص^٢$
١	٥	٥	١	٢٥
٣	٩	٢٧	٩	٨١
٤	٧	٢٨	١٦	٤٩
٥	١	٥	٢٥	١
٧	١٣	٩١	٤٩	١٦٩
٢٠	**٣٥**	**١٥٦**	**١٠٠**	**٣٢٥**

ويلاحظ أن :

$$ع^2_س = \frac{1}{ن - 1} \left[مجـ س^2 - \frac{(مجـ س)^2}{ن} \right] = \frac{1}{4} \left[100 - \frac{(20)^2}{5} \right]$$

$$= \frac{20}{4} = 5 \text{ ، أي أن } ع_س = \sqrt{5} = 2,236$$

كذلك ،

$$ع^2_ص = \frac{1}{ن - 1} \left[مجـ ص^2 - \frac{(مجـ ص)^2}{ن} \right] = \frac{1}{4} \left[325 - \frac{(35)^2}{5} \right]$$

$$= \frac{80}{4} = 20 \text{ ، أي أن } ع_ص = \sqrt{20} = 4,472 \text{ .}$$

كمـا أن :

$$مجـ (س - \bar{س}) (ص - \bar{ص}) = مجـ س ص - \frac{(مجـ س)(مجـ ص)}{ن}$$

$$= 156 - \frac{20 \times 35}{5}$$

$$= 156 - 140 = 16$$

وبالتالي فإن قيمة معامل الارتباط ر تكون :

$$ر = \frac{مجـ (س - \bar{س}) (ص - \bar{ص})}{(ن - 1) ع_س ع_ص}$$

$$= \frac{16}{4 \times 2,236 \times 4,472} = \frac{16}{40,0} = 0,40$$

وهي نفس قيمة المعامل التي حسبت من قبل .

مثال (10) : احسب قيمة معامل الارتباط بين عمر السيـارة المستعملة (س) وسعر بيعها بآلاف الدراهم (ص) باستخدام بيـانات العينـة التاليـة التي

تتألف من ١٢ سيارة بيعت مؤخراً في سوق الحراج .

س : ٣ ٥ ٢ ٦ ٤ ٦ ٣ ١ ٢ ٣ ٤ ٥

ص : ٦ ٤ ٧ ٣ ٣ ٢ ٥ ١٠ ٩ ٤ ٥ ٢

الحل : ينظم العمل الحسابي في الجدول التالي :

س	ص	س ص	$س^٢$	$ص^٢$
٣	٦	١٨	٩	٣٦
٥	٤	٢٠	٢٥	١٦
٢	٧	١٤	٤	٤٩
٦	٣	١٨	٣٦	٩
٤	٣	١٢	١٦	٩
٦	٢	١٢	٣٦	٤
٣	٥	١٥	٩	٢٥
١	١٠	١٠	١	١٠٠
٢	٩	١٨	٤	٨١
٣	٤	١٢	٩	١٦
٤	٥	٢٠	١٦	٢٥
٥	٢	١٠	٢٥	٤
٤٤	**٦٠**	**١٧٩**	**١٩٠**	**٣٧٤**

حيث يلاحظ أن :

$$ع^٢_س = \frac{١}{ن - ١} \left[مجـ س^٢ - \frac{(مجـ س)^٢}{ن} \right]$$

$$= \frac{١}{١١} \left[١٩٠ - \frac{(٤٤)^٢}{١٢} \right] = \frac{١}{١١} [١٩٠ - ١٦١,٣٣]$$

$$= ٢,٧٠$$

أي أن $ع_س = \sqrt{٢,٧٠} = ١,٦٤$

كذلك فإن

$$ع^٢_ص = \frac{١}{ن - ١} \left[مجـ ص^٢ - \frac{(مجـ ص)^٢}{ن} \right]$$

$$= \frac{١}{١١} \left[٣٧٤ - \frac{(٦٠)^٢}{١٢} \right] = \frac{١}{١١} [٣٧٤ - ٣٠٠]$$

$$= ٦,٧٣$$

أي أن $ع_ص = \sqrt{٦,٧٣} = ٢,٥٩$

كما أن :

$$مجـ (س - \bar{س}) (ص - \bar{ص}) = مجـ س ص - \frac{(مجـ س) (مجـ ص)}{ن}$$

$$= ١٧٩ - \frac{٤٤ \times ٦٠}{١٢} = ١٧٩ - ٢٢٠ = -٤١$$

وعلى ذلك فإن قيمة معامل الارتباط ر تكون :

$$ر = \frac{مجـ (س - \bar{س}) (ص - \bar{ص})}{(ن - ١) ع_س ع_ص}$$

$$= \frac{-٤١}{١١ \times ١,٦٤ \times ٢,٥٩} = - \frac{٤١}{٤٦,٧٢} = -٠,٨٨$$

ويعني ذلك وجود علاقة خطية سالبة قوية بين عمر السيارة وسعرها في هذه العينة . ويمكن للقارىء التأكد من ذلك برسم شكل الانتشار المناظر .

ويمكن للقارىء أن يعتمد على أي طريقة من طرق الحساب السابق

شـرحها لإيجـاد قيمة معـامل الارتبـاط . ويتوقف الاختيـار بين هـذه الـطرق المختلفـة على الرأي الشخصي للقـائمين بتحليل البيـانـات ، وتؤدي جميـع الطرق إلى الحصول على الاجابة المطلوبة . وتجـدر الاشارة إلى أنـه يمكن الاستعانة بالبرامج الاحصائية المتوافرة على نطاق واسع للاستخدام على الحاسب الآلي . ذلك أن برنامجاً لحساب معامل الارتباط يتوافر دائماً ضمن جميع مجموعات البرامج الاحصائية المتاحة .

ونختم هذا الفصل عن كيفية حساب قيمة معامل الارتباط بمـلاحظتين هامتين :

أ ـ أن الصيغة المعطاة هي لمعامل الارتبـاط اذا كانت البيـانات تمثـل عينة ، وذلك لأنه من النادر جداً حساب معامل ارتباط بناء على بيانات جميع مفردات المجتمع . ويستخدم معامل الارتباط في العينة لتقدير معامل الارتباط في المجتمـع اعتماداً على أسـاليب الاستنتاج الاحصـائي . ويرمـز لمعـامـل الارتبـاط في المجتمع بـالرمـز اليونـاني ρ ، كمـا أن لـه خصـائص مشـابهـة لخصائص معامل الارتباط في العينة ر .

ب ـ أشرنا في الباب الثالث إلى إمكانية انشاء جداول تكرارية مـزدوجة لوصف نمط العلاقة بين متغيرين س ، ص . وتجدر الإشارة إلى أن الجدول التكراري المزدوج لمتغيرين كميين يعتبر تلخيصـاً لشكل الانتشـار المناظـر . ويمكن حساب قيمة معامل الارتباط بشكل تقـريبي من هذه الجـداول ، وإن كان ذلك أمر نادر الحدوث نتيجة الانتشار الواسع لأساليب الحساب الآلي التي تمكن من رسم شكل الانتشار وحساب قيمة معامل الارتباط باستخدام البيانات الأصليـة ، ودون مجهود حسـابي يذكـر . لذلـك ينصح بـالاكتفاء باستخـدام الجداول التكرارية المزدوجة كأداة للعرض توضح الشكل العـام للعلاقـة بين المتغيرين ، على النحو السابق الاشارة اليه في الباب الثالث .

٧ ـ أوجه القصور في معامل الارتباط

يعجز معامل الارتباط في بعض المواقف عن اعطاء وصف مناسب لقوة العلاقة بين متغيرين . وفيما يلي عرض موجز لأهم هذه المواقف .

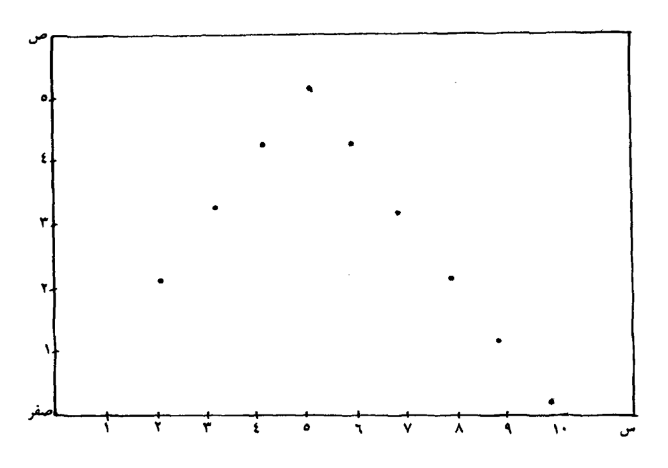
أ ـ العلاقة بين المتغيرين غير خطية : يقيس معامل الارتباط درجة تركز نقط شكل الانتشار حول خط مستقيم ، ولذلك يسمى عادة معامل الارتباط الخطي . اذا كانت العلاقة بين المتغيرين غير خطية فان معامل الارتباط يكون مقياساً غير جيد لهذه العلاقة ، كما يتضح من المثال التالي :

مثال (١١) : احسب معامل الارتباط الخطي بين س ، ص للبيانات التالية ، ثم ارسم شكل الانتشار لهذه البيانات وعلق على مدى ملاءمة معامل الارتباط لوصف قوة العلاقة بين المتغيرين .

س :	١٠	٢	٩	٧	٣	٨	٥	٤	٦
ص :	صفر	٢	١	٣	٣	٢	٥	٤	٤

الحل : قيمة معامل الارتباط الخطي في هذه الحالة هي ر = ـ١١و٠ ويترك للقارىء حساب هذه القيمة كتمرين . يعطي شكل (١٨) شكل الانتشار المناظر ، حيث يلاحظ وجود علاقة غير خطية تامة بين المتغيرين س، ص ، وذلك على الرغم من صغر قيمة معامل الارتباط . ويدل ذلك على ان معامل الارتباط الخطي لا يصلح لقياس قوة العلاقات غير الخطية . لذلك يكون من الضروري دائماً دراسة شكل الانتشار عند محاولة تفسير معنى قيمة معامل الارتباط . ذلك ان صغر قيمة معامل الارتباط الخطي لا يدل على ضعف العلاقة بين المتغيرين بشكل عام ، وانما يدل فقط على ضعف العلاقة الخطية بينهما .

وتجدر الاشارة الى وجود أساليب احصائية خاصة لمعالجة مثل هذه المواقف ، اذ يمكن مثلاً استخدام تحويلة مناسبة للحصول على علاقة خطية في البيانات . وقد سبقت مناقشة مفهوم التحويلات بإيجاز في الباب الرابع .



شكل (١٨) : شكل الانتشار للعلاقة بين س ، ص

ب ـ وجود مشاهدات شاذة أو متطرفة في البيانات : يتأثر معامل الارتباط بوجود مشاهدات متطرفة في البيانات اذ قد يؤدي الاعتماد على معامل الارتباط لقياس قوة العلاقة بين متغيرين في هذه الحالات الى نتائج مضللة كما يتضح من المثال التالي .

مثال (١٢) : فيما يلي بيانات عن متغيرين س ، ص لعينة من ٥ مفردات :

س :	١	٢	٣	٤	٥
ص :	١	٢	٣	٤	صفر

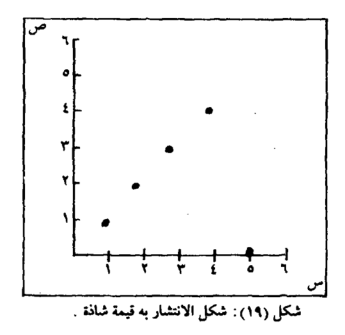
يوضح شكل (١٩) شكل الانتشار المناظر . يلاحظ ان المشاهدة الاخيرة (٥، ٠) تختلف بشكل واضح عن بقية المشاهدات . اذا حسبت قيمة معامل الارتباط باستخدام جميع المشاهدات فان قيمة ر تساوي صفراً وذلك

على الـرغم من وجـود نمط خـطي للعـلاقــة بين س، ص . إذا استبعـدت المشاهدة (٥، ٠) وحسبت قيمـة معامـل الارتباط على أسـاس المشاهـدات الباقية فان هذه القيمة تساوي الواحد .

يتضح من ذلك ان وجود مشاهدات متطرفـة في البيانـات قد يؤدي الى لبس عند استخدام معـامل الارتبـاط كمقياس وصفي . ولـذلك ينصـح دائماً بالنظر الى شكـل الانتشار قبـل الارتكان الى معـامل الارتبـاط كمقياس لقـوة العلاقة بين المتغيرين .

وقد سبقت الاشارة الى ان أسلوب معالجة المشـاهدات الشـاذة يعتمد على طبيعة البيانات وظروف الدراسة . وعـادة ما يتم عـزل هذه المشـاهدات ودراستها بصفة مستقلة ، أو اهمالها تماماً اذا كان هناك من الأسباب ما يشجع على ذلك .

جـ ـ قياس الارتباط بين المتوسطات : افتـرض انه يـراد حساب قيمـة معامل الارتباط بين مستوى الدخل ومستوى التعليم للذكور في قوة العمل في دولة ما . قام احد الباحثين بجمع بيانات عن الدخل والتعليم لهؤلاء الذكور من تعداد السكان وحسب قيمة المعامل المطلوب على أساسها . لاحظ باحث آخر أن الدولة مقسمة الى عشر مناطق جغرافية ، فقرر تبعاً لذلك حساب متـوسط مستـوى التعليم ومتوسط مستوى الدخل في كل منطقة ، ثم استخـدم هذه المشـاهدات العشـر لحسـاب معـامـل ارتبـاط بين مستـوى التعليم ومستـوى الدخل . هل يحصل الباحثان على نفس قيمة معامل الارتباط ؟



شكل (١٩) : شكل الانتشار به قيمة شاذة .

تكـون الإجابـة عن مثل هـذا السؤال عادة بـالنفي ، إذ يؤدي استخدام المتوسطات كأساس لحساب معامل الارتباط الى نتائج مضللة ، كما يتضح من المثال التالي :

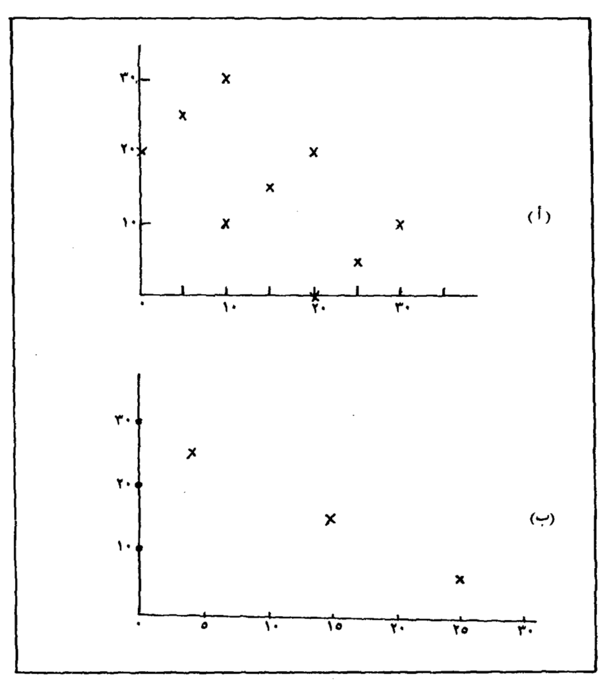
مثال (١٣) : هناك بيـانات عن المتغيـرين س، ص لثلاث مجمـوعات متميزة من المشاهدات ؛ هذه البيانات هي :

المجموعة الأولى :	س :	٠	٥	١٠
	ص :	٢٠	٢٥	٣٠
المجموعة الثانية :	س :	١٠	١٥	٢٠
	ص :	١٠	١٥	٢٠
المجموعة الثالثة :	س :	٢٠	٢٥	٣٠
	ص :	٠	٥	١٠

ويظهر شكل الانتشار المناظر لهـذه البيانـات في شكل ٢٠ (أ) . اذا حسبت قيمة معامل الارتباط على أساس جميع المشاهدات في شكل الانتشار فإن الناتج يساوي ـ٠,٦٠ ، مما يدل على وجود علاقة خطيـة سالبـة ومتوسـطة القوة. أما إذا أخذت متوسطات قيم س وقيم ص كأسـاس لحساب الارتبـاط، فإن شكل الانتشار المناظـر يظهـر في شكل (٢٠ ـ ب) . ويـلاحظ ان قيمة معامل الارتباط في هذه الحالة تساوي ـ١ ، مما يدل على وجود علاقة خـطية سـالبة تـامة . ويعني ذلـك انه لا يمكن الاعتمـاد على معامـل الارتبـاط بين المتـوسطات كبـديل لمعـامل الارتبـاط في المشاهـدات الأصليـة . ويـرجـع الاختلاف في قيمة المعاملين الى وجود تشتت لمشاهدات كل مجموعة حول متـوسـطهـا . هـذا التشتت لا يؤخـذ في الاعتبـار عنـد الاكتفـاء بـاستخـدام المتوسطات كأساس للحساب . ويؤدي ذلك الى اعطاء الانطباع الخادع بـأن نقط الشكل اكثر تـركزاً حـول خط مستقيم ، كما يتضـح من القيمة المـرتفعة لمعامل الارتباط في هذه الحالة . قارن بين شكلي ٢٠ (أ) ، ٢٠ (ب) .

وينبغي ان يتـذكر القـارىء ، كملاحظ أخيـرة ، أن معامـل الارتباط لا

يكفي بمفرده لوصف شكل الانتشار . اذ يجب التأكيد على أن هذا المعامل يستخدم بالاضافة الى المتوسطات والانحرافات المعيارية للمتغيرات تحت الدراسة . وتقدم المقاييس الخمس معا معلومات جيدة تكون كافية في معظم الحالات لوصف السمات الأساسية في مجموعات البيانات المزدوجة .



شكل (٢٠)

٨ ـ معامل ارتباط الرتب

يعاب على معامل الارتباط الخطي تأثره بوجود القيم الشاذة أو المتطرفة في البيانات ، هذا بالاضافة الى عدم صلاحيته لقياس العلاقات غير الخطيـة بين المتغيرات . ويمكن الاعتماد في مثل هذه المـواقف على معامـل ارتباط الرتب كبديل لوصف العلاقة بين متغيرين س،ص . ذلك ان معـامل ارتبـاط الرتب يكون أقل تأثراً في حالة ابتعاد العلاقة عن الشكل الخطي ، فضلاً عن كونه أقل حساسية للقيم الشاذة أو المتطرفة التي قد توجد في البيانات .

ويعتمـد معـامـل ارتبـاط الـرتب على رتب س،ص في المشـاهــدات المختلفة بدلاً من الاعتمـاد على القيم الفعلية لهـذه المشاهـدات . وتحسب قيمـة هذا المعـامل بتـرتيب قيم س ترتيبـاً تصاعـدياً وتـرتيب قيم ص تـرتيبـاً تصـاعديـاً ، ثم حساب معـامل الارتبـاط الخطي بين رتب قيم س ورتب قيم ص .

مثال (١٤): احسب قيمة معامل ارتباط الرتب بين س، ص بـاستخدام البيانات الآتية عن عينة من ٩ مفردات .

س :	٢٢١	٢٢٨	٢٢٣	٢١١	٢٣١	٢١٥	٢٢٤	٢٣٣	٢٦٨
ص :	٠,٦٧	٠,٨٦	٠,٧٨	٠,٥٤	٠,٩١	٠,٤٤	٠,٩٠	٠,٩٤	٠,٩٣

الحل : نبدأ بترتيب قيم س تصاعدياً على الشكل :

٢١١ ، ٢١٥ ، ٢٢١ ، ٢٢٣ ، ٢٢٤ ، ٢٢٨ ، ٢٣١ ، ٢٣٣ ، ٢٦٨

وعلى ذلك فان رتبة أول قيمة من قيم س (أي ٢٢١) تسـاوي ٣ ورتبة ثـاني قيمة (أي ٢٢٨) تساوي ٦ وهكذا تكون رتب قيم س هي ٣، ٦، ٤، ١، ٧، ٢، ٥، ٨، ٩ .

نرتب قيم ص تصاعدياً أيضاً على الشكل :

٤٤، ٠,٥٤، ٠,٦٧، ٠,٧٨، ٠,٨٦، ٠,٩٠، ٠,٩١، ٠,٩٣، ٠,٩٤

أي أن رتب قيم ص هي : ٣، ٥، ٤، ٢، ٧، ١، ٦، ٩، ٨ .

ثم بعد ذلك تحسب معامل الارتباط الخطي بين رتب قيم س ورتب قيم ص . وتظهر خطوات الحساب في الجدول التالي :

س	ص	رتب س	رتب ص	رتب س×رتب ص	(رتب س)٢	(رتب ص)٢
٢٢١	٠,٦٧	٣	٣	٩	٩	٩
٢٢٨	٠,٨٦	٦	٥	٣٠	٣٦	٢٥
٢٢٣	٠,٧٨	٤	٤	١٦	١٦	١٦
٢١١	٠,٥٤	١	٢	٢	١	٤
٢٣١	٠,٩١	٧	٧	٤٩	٤٩	٤٩
٢١٥	٠,٤٤	٢	١	٢	٤	١
٢٢٤	٠,٩٠	٥	٦	٣٠	٢٥	٣٦
٢٣٣	٠,٩٤	٨	٩	٧٢	٦٤	٨١
٢٦٨	٠,٩٣	٩	٨	٧٢	٨١	٦٤
		٤٥	٤٥	٢٨٢	٢٨٥	٢٨٥

وتكون صيغة حساب معامل ارتباط الرتب على الشكل التالي :

$$\text{معامل ارتباط الرتب} = \frac{\text{مجـ (رتبةس ـ متوسط رتب س) (رتبة ص ـ متوسط رتبة ص)}}{\text{(ن ـ ١) . الانحراف المعياري لرتب س} \times \text{الانحراف المعياري لرتب ص}}$$

وباتباع نفس الأساليب السابقة ، نلاحظ أن :

مجـ(رتبة س ـ متوسط رتب س) (رتبة ص ـ متوسط رتب ص) =

$$\text{مجـ رتبة س} \times \text{رتبة ص} - \frac{\text{(مجـ رتب س) (مجـ رتبة ص)}}{\text{ن}}$$

$$= ٢٨٢ - \frac{٤٥ \times ٤٥}{٩}$$

$$= ٢٨٢ - ٢٢٥ = ٥٧$$

كما أن الانحراف المعياري لرتب س $= \sqrt{\frac{١}{٨}\left[٢٨٥ - \frac{(٤٥)^{٢}}{٩} \right]}$

$$= \sqrt{٧,٥} = ٢,٧٣٩$$

والانحراف المعياري لرتب ص $= \sqrt{\frac{١}{٨}\left[٢٨٥ - \frac{(٤٥)^{٢}}{٩} \right]} = \sqrt{٧,٥}$

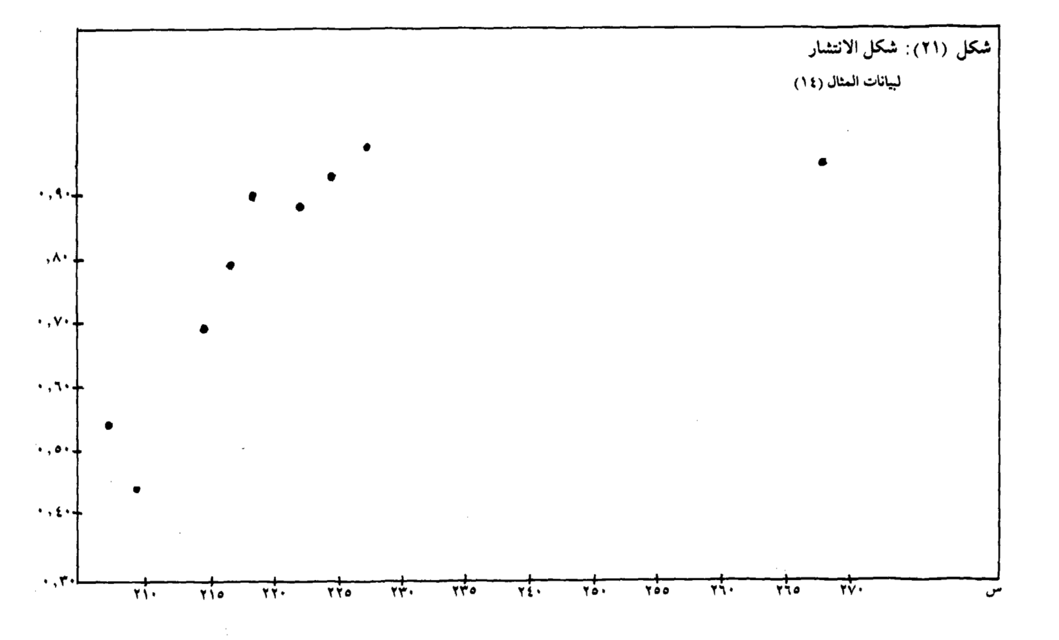
$$= ٢,٧٣٩$$

وتكون قيمة معامل ارتباط الرتب $= \frac{٥٧}{٨ \times ٢,٧٣٩ \times ٢,٧٣٩} = \frac{٥٧}{٦٠} = ٠,٩٥$

وتدل هذه القيمة على وجود علاقة قوية بين س، ص . ذلـك ان قيمة معامل ارتباط الرتب تتراوح بين ـ١ ، ١ بحيث انه كلما اقتربت قيمة المعامـل من الواحد كلما دل ذلك على قوة العلاقة بين المتغيرين ، وكلما اقتربت قيمة المعامل من الصفر كلما دل ذلك على ضعف العلاقة . وتكون قيمة المعامل مساوية ١ اذا كانت كل قيمة كبيرة للمتغير س يناظرها قيمة كبيرة للمتغير ص أو العكس ، وتكون قيمة المعامل مساوية ـ١ اذا كانت كل قيمة كبيرة للمتغير س يناظرها قيمة صغيرة للمتغير ص أو العكس .

يعطي شكل (٢١) شكل الانتشار المناظر لبيـانات مثـال (١٤) ، حيث يلاحظ ان هناك مشاهدة متطرفة هي (٢٦٨ ، ٠,٩٣) . وقد أثر ذلك في قيمة معامل الارتباط الخطي التي تساوي ٠,٦٦ في هذه الحالة . (يتـرك حساب قيمة هذا المعامل كتمرين) . اذا استبعدت هذه المشاهـدة من البيانـات فان قيمة معامل الارتباط الخطي تقفز الى ٠,٩١ بينما تظل قيمة معامل ارتباط الرتب عند القيمة ٠,٩٥ (ويمكن للقارىء التأكد من صحة هذه النتائج) ، ويوضح ذلك ان معامل ارتباط الـرتب أقل تـأثراً بـالقيم المتطرفـة من معامـل الارتباط الخطي .

وتجـدر الاشارة الى ان معـامل ارتبـاط الرتب يصلح لـلاستخـدام مـع



شكل (٢١): شكل الانتشار
لبيانات المثال (١٤)

البيانات الترتيبية . فمثلاً اذا طلب من شخصين ترتيب عشر برامج تليفزيونية حسب تفضيلهم لها ، فان معامل ارتباط الرتب يمكن ان يستخدم لقياس درجة الاتفـاق بين آراء الشخصين . ولا يكون هنـاك داع في هذه الحـالـة لتـرتيب البيانات ، لأن هذه البيانات معطاة في صورة مرتبة أصلاً .

تمرينات

١ - استخدم شكل (٨) الذي يمثل شكل الانتشار لطول الأب وطول الابن للاجابة عن الأسئلة الآتية :

(أ) ما هو طول أقصر أب في العينة ؟ وما هو طول إبنه ؟

(ب) ما هو طول أقصر ابن وما هو طول أطول ابن للآباء الذين يبلغ طولهم ٧٢ بوصة ؟

(حـ) ما هو عدد الأسر التي يبلغ طول الابناء فيها ٧٦ بوصة ؟

(د) ما هي القيمة التقريبية لمتوسط طول الآباء ؟

(هـ) ما هي القيمة التقريبية للانحراف المعياري لطول الآباء ؟

٢ - في دراسة عن العلاقة بين طول الشخص ووزنه في عينة من طلبة الجامعة ، وجد أن متوسط طول الشخص يساوي ١٧٠سم وان الانحراف المعياري للطول = ٨سم . كذلك وجد ان متوسط وزن الشخص يساوي ٦٣كجم وأن الانحراف المعياري للوزن = ٩كجم . كذلك كانت قيمة معامل الارتباط بين الطول والوزن تساوي ٠,٦٠

(أ) اذا كان طول أحد الطلبة مساوياً ١٨٢سم ، ماذا يجب ان يكون عليه وزنه حتى يقع الطالب على خط الارتباط التام ؟

(ب) هل تقع النقطة (١٧٨سم ، ٧٢كجم) على خط الارتباط التام ؟

٣ - اذا كانت قيمة معامل الارتباط بين الطول والوزن لمجموعة من الذكور تساوي ٠,٧ وكانت قيمة هذا المعامل لمجموعة من الإناث تساوي أيضاً ٠,٧ اذا حسبت قيمة معامل الارتباط باستخدام بيانات الذكور والاناث معاً فهل تختلف هذه القيمة عن ٠,٧ ؟ولماذا ؟

٤ : في دراسة عن العلاقة بين مستوى ذكاء الزوج ومستوى ذكاء الزوجة وجد أن :

متوسط درجة ذكاء الزوج = ١٠٠ والانحراف المعياري = ١٥

وأن متوسط درجة ذكاء الزوجة = ١٠٠ والانحراف المعياري = ١٥
وأن قيمة معامل الارتباط = ٠,٦
ارسم شكل انتشار لهذه العلاقة بحيث يتفق شكله العام مع هذه النتائج .

٥ - يعطي شكل (٢٢) مجموعة أشكال انتشار مختلفة . حدد شكل الانتشار الذي يناظر كل من قيم معامل الارتباط التالية : - ٠,٨٥ ؟ - ٠,٣٨ ؟ - ١ ؟ + ٠,٠٦ ؟ + ٠,٦٢ ؟ ٠,٩٧

٦ - (أ) علق على مدى صحة العبارة التالية : « إذا كانت قيمة معامل الارتباط الخطي تساوي صفراً ، فان ذلك يعني عدم وجود علاقة بين المتغيرين » .

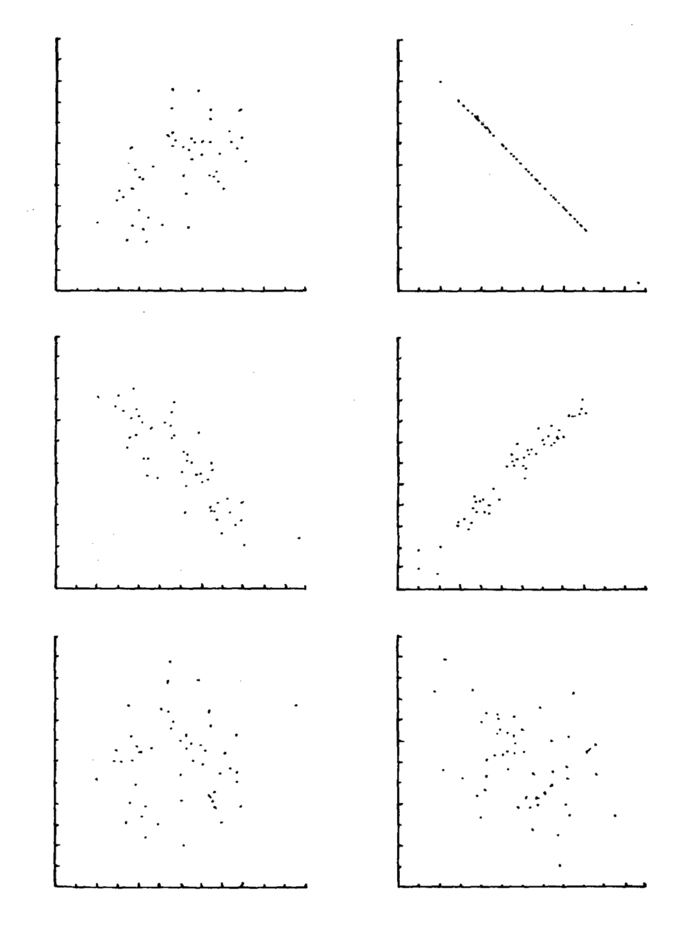
(ب) اذا كان عمر الزوج يزيد دائماً عن عمر زوجته بمقدار خمس سنوات فما هي قيمة معامل الارتباط بين عمر الزوج وعمر الزوجة ؟ اشرح سبب اجابتك .

(حـ) كان مطلوباً في احد الاختبارات الاجابة عن عشر أسئلة . قام مدرس المساق بحصر عدد الاجابات الصحيحة وعدد الاجابات الخاطئة لكل طالب . وجد أن متوسط عدد الاجابات الصحيحة يساوي ٧,٢ ومتوسط عدد الاجابات الخاطئة يساوي ٢,٨ . وكانت قيمة الانحراف المعياري لعدد الإجابات الصحيحة تساوي ٢ وقيمة الانحراف المعياري لعدد الاجابات الخاطئة يساوي ٢ أيضاً .

ما هي قيمة معامل الارتباط بين عدد الاجابات الصحيحة وعدد الاجابات الخاطئة ؟

٧ - (أ) ما هي قيمة معامل الارتباط بين س ، ص في البيانات الآتية ، ولماذا ؟

س :	١	١	١	١	٢	٢	٢	٣	٣	٤
ص :	٢	٢	٢	٢	٤	٤	٤	٦	٦	٨



شكل (٢٢)

(ب) احسب قيمة معامل الارتباط بين س، ص في البيانات الآتية :

س : ١ ١ ١ ١ ٢ ٢ ٢ ٣ ٣ ٤

ص : ١ ٢ ١ ٣ ١ ٤ ١ ٢ ٢ ٣

(حـ) هل تتغير قيمة معامل الارتباط بين س ، ص في (ب) اذا استبدلت قيم س بقيم ص واستبدلت قيم ص بقيم س ؟

(د) هل تتغير قيمة الارتباط في (ب) اذا ضربت كل قيمة من قيم س في ٢ وأضيفت القيمة ١٥ الى كل قيمة من قيم ص ؟

٨ ـ تعطي البيانات التالية قيم مقياسين مختلفين لاستهلاك الوقود لسبعة أنواع مختلفة من السيارات . يمثل المقياس الأول عدد الأميال التي تقطعها كل سيارة بجالون واحد من الوقود ، بينما يمثل المقياس الثاني عدد الليترات اللازمة لقطع مسافة ١٠٠كم .

نوع السيارة	عدد الأميال لكل جالون	عدد الليترات لكل ١٠٠كم
أ	٢٣	١٠,٣
ب	٣٥	٦,٨
جـ	١٩	١٢,٤
د	٢٧	٨,٨
هـ	٤١	٥,٨
و	١٨	١٣,١
ز	١٢	١٩,٧

(أ) احسب قيمة معامل الارتباط بين المقياسين . ارسم شكل الانتشار وعلق على مدى ملاءمة معامل الارتباط الخطي لوصف العلاقة بينهما .

(ب) اذا كان المقياس الثاني هو عدد الليترات لكل ١٠٠٠كم (بدلاً من ١٠٠كم) فهل تتغير قيمة معامل الارتباط ؟ اشرح سبب الاجابة .

٩ ـ (أ) تظهر العديد من الدراسات أن هناك ارتباطاً بين التدخين وأمراض

القلب . كذلك ، اظهرت احدى الدراسات الحديثة ان هناك ارتباطاً بين شرب القهوة وأمراض القلب هل يستنتج من ذلك أن شرب القهوة يسبب أمراض القلب ؟ اشرح سبب اجابتك .

(ب) لوحظ في احدى الدراسات وجود ارتباط موجب بين مستوى دخل الشخص وارتفاع ضغط دمه . هل يستنتج من ذلك ان ارتفاع الدخل يؤدي الى ارتفاع ضغط الدم ؟ اشرح سبب الاجابة .

(جـ) يرغب أحد الباحثين في دراسة العلاقة بين جنسية الأب وعدد أفراد أسرته . هل يمكن استخدام معامل الارتباط الخطي لوصف هذه العلاقة ؟ وضح سبب اجابتك .

١٠ ـ أعطيت البيانات التالية في احدى الدراسات عن العلاقة بين التدخين ومستوى الوفاة . تمثل البيانات متوسط نصيب الفرد من عدد السجائر المستهلكة في الدولة ومعدل الوفاة من سرطان الرئة (للمليون) في الدولة وذلك في عام ١٩٥٠ .

الدولة	متوسط استهلاك السجائر للفرد	معدل الوفاة (للمليون) من سرطان الرئة
استراليا	٤٨٠	١٨٠
كندا	٥٠٠	١٥٠
الدانمارك	٣٨٠	١٧٠
فنلندا	١١٠٠	٣٥٠
بريطانيا	١١٠٠	٤٦٠
هولندا	٤٩٠	٢٤٠
ايسلندا	٢٣٠	٦٠
النرويج	٢٥٠	٩٠
السويد	٣٠٠	١١٠
سويسرا	٥١٠	٢٥٠
أمريكا	١٣٠٠	٢٠٠

(أ) ارسم شكل الانتشار لهذه البيانات .

(ب) هل يظهر شكل الانتشار علاقة بين مستوى استهلاك السجائر ومعدل الوفاة لسرطان الرئة في الدول المختلفة ؟

(جـ) هل يظهر شكل الانتشار علاقة بين تدخين الشخص ووفاته بسرطان الرئة ؟ اشرح سبب اجابتك .

١١ ـ علق على مدى صحة العبارات الآتية :

(أ) « إذا كانت قيمة معامل الارتباط تساوي ـ ٠,٨٥ فان ذلك يعني ان الغالبية العظمى من النقط في شكل الانتشار تتركز في الجزء اليساري الأسفل والجزء اليميني الأعلى من شكل الانتشار » .

(ب) « إذا كانت قيمة المتغير المستقل تقل عادة عن قيمة المتغير التابع فان ذلك يؤدي الى قيمة سالبة لمعامل الارتباط » .

١٢ ـ بالرجوع الى بيانات التمرين رقم (١٤) صفحة (٣١٩) ، يلاحظ وجود خطأ واضح في مشاهدات طالبين .

(أ) ارسم شكل الانتشار لهذه البيانات وعلق على نمط العلاقة بين القياس في المرة الأولى والقياس في المرة الثانية .

(ب) احسب قيمة معامل ارتباط الرتب في هذه الحالة .

(حـ) استبعد المشاهدتين الشاذتين وأعد حساب قيمة معامل ارتباط الرتب للمشاهدات الباقية . قارن بالاجابة في (ب) .

١٣ ـ في كل حالة من الحالات التالية ، هل تتوقع ان تكون قيمة معامل الارتباط موجبة ام سالبة أم صفرية ؟ اشرح سبب اجابتك :

(أ) العلاقة بين درجة الحرارة اثناء النهار وكمية الكهرباء المستهلكة .

(ب) العلاقة بين طول الشخص ومستوى ذكائه .

(حـ) العلاقة بين طول الشخص ومقاس حذائه .

(د) العلاقة بين الزمن الذي يقضيه الطفل في مشاهدة برامج التلفزيون يومياً والزمن الذي يقضيه في اتمام واجبه المدرسي .

١٤ ـ في دراسة عن العلاقة بين دخل الأسرة (س) وما تنفقه على الطعام والشراب (ص) في عينة من ٢٤ أسرة في احدى البلدان وجد أن : ن = ٢٤ ، ع‌س = ٨٠٦,٥٤ ، ع‌ص = ٨٩,١٢ ، مجـ (س - سَ) (ص - صَ) = ١٦٤٧,٣٠ . احسب قيمة معامل الارتباط وفسر معناها .

١٥ ـ احسب قيمة معامل الارتباط بين س، ص اذا علم أن :

ن = ٢٣ ، سَ = ١٤٢,٥ ، مجـ س^{٢} = ١٠٢٤٦٢٥,٨٣

صَ = ٤٧٣,٠ مجـ ص^{٢} = ١١٤٩٢٩٥٧,٠١

مجـ س ص = ٢٩٦١١٩٠,٥٨

فسر معنى القيمة التي تحصل عليها .

١٦ ـ تعطي البيانات التالية قيمة مقياس التغير في أسعار الوقود وقيمة مقياس التغير في أسعار السلع والخدمات الأخرى في بلد ما خلال الفترة الزمنية ١٩٧٠ ـ ١٩٨١ :

السنة :	١٩٧٠	١٩٧١	١٩٧٢	١٩٧٣	١٩٧٤	١٩٧٥
مقياس أسعار الوقود :	٥١,٧	٥٣,٧	٥٣,٨	٥٥,٧	٧٩,٢	٨٩,٦
مقياس أسعار السلع والخدمات :	٦٤,٠	٦٦,٨	٦٩,٠	٧٣,٣	٨١,٤	٨٨,٨

	١٩٧٦	١٩٧٧	١٩٧٨	١٩٧٩	١٩٨٠	١٩٨١
	٩٧,٦	١٠٠,٠٠	١٠٦,٦	١٥٣,٣	٢٠٦,٢	٢٢٩,٩
	٩٣,٩	١٠٠,٠٠	١٠٧,٧	١١٩,٨	١٣٦,٠	١٥٠,١

(أ) ارسم شكل الانتشار وعلق على نمط العلاقة في البيانات .

(ب) احسب قيمة معامل الارتباط الخطي بين مقياس أسعار الوقود ومقياس أسعار السلع والخدمات الأخرى .

(جـ) احسب قيمة معامل ارتباط الرتب .

(د) اشرح السبب الذي يجعل قيمة معامل ارتباط الرتب مساوية ١ بينما قيمة معامل الارتباط الخطي تقل عن الواحد .

١٧ ـ في دراسة عن العوامل المختلفة التي ترتبط بمستوى الجريمة في المناطق المختلفة ، رتبت سبع مناطق حسب مستوى الجريمة ، وحسب نسبة العمال الأجانب ، وحسب قيمة وسيط الدخل ، وحسب نسبة البطالة في كل منها ، فكانت البيانات التالية :

المنطقة	ترتيب مستوى الجريمة	ترتيب نسبة العمال الأجانب	ترتيب قيمة وسيط الدخل	ترتيب نسبة البطالة
أ	٢	٣	٧	١
ب	٤	٦	٦	٤
جـ	٧	٧	١	٧
د	٥	٥	٢	٥
هـ	٦	٤	٣	٦
و	٣	٢	٥	٣
ز	١	١	٤	٢

(أ) احسب معامل ارتباط الرتب بين مستوى الجريمة ونسبة العمال الأجانب .

(ب) احسب معامل ارتباط الرتب بين مستوى الجريمة وقيمة وسيط الدخل .

(جـ) احسب معامل ارتباط الرتب بين مستوى الجريمة ونسبة البطالة .

(د) حدد بناءاً على النتائج السابقة أي المتغيرات اكثر ارتباطاً بمستوى الجريمة .

١٨ ـ يقوم شخصان بتذوق ٩ أنواع من الشاي ، بحيث يقوم كل منهم بترتيب هـذه الأنواع حسب أفضليته لها . فيما يلي البيانات الخاصة بهـذه التجربة :

نوع الشاي :	أ	ب	جـ	د	هـ	و	ز	ل	ك
الرتبة التي أعطاها الشخص الأول :	٧	١	٣	٢	٨	٥	٩	٦	٤
الرتبة التي أعطاها الشخص الثاني :	٩	٤	١	٣	٧	٥	٦	٨	٢

احسب قيمة معامل ارتباط الرتب كمقياس لدرجة الاتفاق في آراء الشخصين .

١٩ ـ (أ) افترض ان المتغيرين س، ص يأخذان قيماً موجبة وأن معامل الارتباط المحسوب من عينة حجمها ىر من المشاهدات يساوي الواحد . اذا ربعت قيم ص ، هل تظل قيمة معامل الارتباط بين س ، ص^{٢} مساوية للواحد ؟ اشرح سبب اجابتك .

(ب) في الجزء (أ) ، اذا كانت قيمة معامل ارتباط الرتب بين س ، ص يساوي الواحد ، هل تظل قيمة هذا المعامل بين س ، ص^{٢} مساوية للواحد ؟ اشرح سبب اجابتك .

٢٠ ـ هناك صيغة رياضية أخرى لحساب قيمة معامل ارتباط الرتب . اذا كانت ف تمثل الفرق بين رتبة س والرتبة المناظرة للمتغير ص فإن قيمة معامل ارتباط الرتب يمكن ان تحسب بالعلاقة :

$$\text{معامل ارتباط الرتب} = ١ - \frac{٦ \text{ مجـ ف}^{٢}}{\text{ىر} (\text{ىر}^{٢} - ١)}$$

(أ) استخدم هذه الصيغة لحساب معامل ارتباط الرتب للبيانات في التمرين (١٨) .

(ب) استخدم هذه الصيغة لبيان أن قيمة معامل ارتباط الرتب تساوي الواحد إذا كان هناك ارتباط تام موجب بين المتغيرين س ، ص .

الباب العاشر

الانحدار

لا يقتصر الأمر عند دراسة العلاقة بين متغيرين على حساب قيمة معامل الارتباط بينهما ، بل يتعداه بالضرورة الى محاولة الحصول على وصف كمي لشكل هذه العلاقة باستخدام معادلة رياضية مناسبة. وتفيد مثل هذه المعادلة في تلخيص خصائص وسمات شكل الانتشار ، هذا فضلاً عن امكانية استخدامها لأغراض التنبؤ بقيم أحد المتغيرات اعتماداً على قيم المتغير الآخر .

يهتم أسلوب الانحدار بدراسة كيفية التغير في قيمة المتغير التابع (ص) مع تغير المتغير المستقل (س) وتحديد معادلة رياضية لوصف العلاقة بين المتغيرين ، ثم تقويم كفاءة هذه المعادلة ودراسة نمط تشتت نقط شكل الانتشار حولها . وتجدر الاشارة الى ان مفهوم التنبؤ الاحصائي (بمعنى الاعتماد على أحد المتغيرين لتقدير قيم المتغير الآخر) يعتبر مفهوماً أساسياً في علاقات الانحدار . وعلى ذلك يكون من الضروري تحديد المتغير المستقل (الذي يتخذ أساساً للتنبؤ) والمتغير التابع (الذي يراد التنبؤ بقيمته) كخطوة أولى عند إجراء الدراسة . وقد يكون هذا التحديد واضحاً ومنطقياً في بعض المواقف ، كما قد يكون غير واضح في مواقف أخرى . فمثلاً ، عند دراسة العلاقة بين عمر السيارة وسعر بيعها ، يكون من المنطقي اعتبار عمر السيارة متغيراً مستقلاً واعتبار سعر بيعها متغيراً تابعاً ؛ اذ يمكن الاعتماد على عمر السيارة كأساس للتنبؤ بسعر بيعها . أما إذا كانت هناك بيانات عن كمية الطعام التي يستهلكها مجموعة من

الأشخاص السمان يومياً وعن أوزان هؤلاء الأشخاص ، فانه قد يراد الاعتماد على كمية الطعام المستهلكة للتنبؤ بوزن الشخص أو العكس، قد يؤخذ وزن الشخص كأساس للتنبؤ بكمية الطعام التي يستهلكها . ولا يمكن للتحليل الاحصائي في مثل هذه الحالات ان يتم دون تحديد واضح للهدف من اجراء الدراسة يستخدم في تصنيف المتغيرات الى متغيرات مستقلة واخرى تابعة . ونناقش في الأجزاء التالية تفاصيل أسلوب الانحدار في دراسة العلاقة بين متغيرين .

١ ـ مثال أولي

يبحث أحد رجال الأعمال عن مكان لافتتاح سوق مركزي لخدمة سكان احدى المدن . ومن المرغوب فيه ان يكون مكان هذا السوق قريب من مركز المدينة بقدر الامكان ، وقد يتطلب ذلك ارتفاعاً في تكلفة الايجار . لذلك قام رجل الأعمال بدراسة العلاقة بين تكلفة ايجار المكان وبعده عن مركز المدينة كخطوة اولى في بحث جدوى المشروع .

يعطي جدول (١) بيانات جمعتها غرفة التجارة والصناعة في المدينة عن ايجارات عشرة اماكن مختلفة وعن بعد كل منها عن مركز المدينة ، ويراد مناقشة كيفية الاستفادة من هذه البيانات لأغراض التنبؤ بتكلفة الايجار للأماكن المختلفة . افترض ان هناك مكاناً يبعد عن مركز المدينة مسافة ٤٠كم ويراد تخمين تكلفة ايجاره . قد يكون من المنطقي في هذه الحالة الاعتماد على متوسط تكلفة الايجار لجميع الأماكن المعطاة في العينة كتخمين أولي ، وتكون الاجابة المطلوبة هي ٥٧ درهماً للمتر المربع . وتجدر الاشارة الى ان الاعتماد على المتوسط كأساس للتخمين والتنبؤ هو امر منطقي ومقبول في الحالات التي لا تتوافر فيها معلومات اضافية قد تبعث على الاعتقاد بأن التقدير المطلوب يختلف عن المتوسط بشكل أساسي .

جدول (١)

بيانات عن تكلفة الايجار والبعد عن مركز المدينة لعشرة أماكن مختلفة

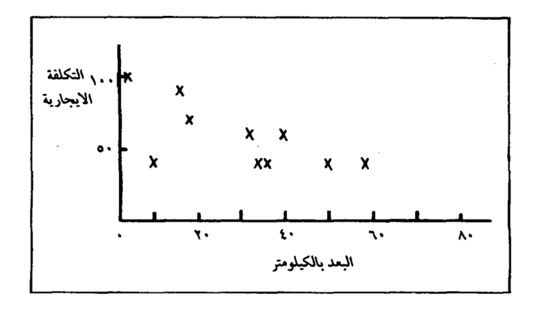
المكان	تكلفة الايجار للمتر المربع بالدرهم	البعد عن مركز المدينة بالكيلومتر
١	٩٧	٤
٢	٣٥	١٠
٣	٨٨	١٧
٤	٧٠	١٩
٥	٦٣	٣١
٦	٣٦	٣٤
٧	٤٢	٣٦
٨	٦٢	٤١
٩	٣٥	٤٩
١٠	٤١	٥٨

لا يتوقع بالطبع ان تكون قيمة التنبؤ الناتجة وهي ٥٧ درهماً للمتر المربع دقيقة تماماً وان كان من المؤمل ان تكون قريبة من القيمة الحقيقية بقدر الامكان . ويمكن قياس درجة دقة هذا التنبؤ باتباع نفس الأسلوب لتقدير تكلفة ايجار كل مكان من الأماكن العشرة المعطاة ثم حساب قيمة الخطأ في كل حالة . وتكون هذه الأخطاء هي :

٤٠ ، ـ ٢٢ ، ٣١ ، ١٣ ، ٦ ، ـ٢١ ، ـ ١٥ ، ٥ ، ـ ٢٢ ، ـ ١٦ ،

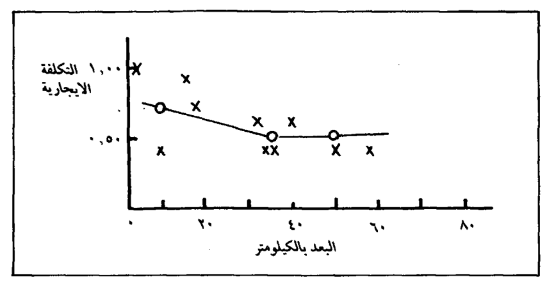
وهي أخطاء ذات حجم كبير اذ يبلغ متوسطها بصرف النظر عن الإشارة أكثر من ١٩ درهم . ويعني ذلك عدم دقة الأسلوب المتبع في التخمين ، والحاجة الى أساليب أخرى أكثر دقة للحصول على التقدير المطلوب . ويعتبر الانحدار أحد أهم هذه الأساليب .

يعتمد أسلوب الانحدار في التخمين والتنبؤ على جمع بيانات اضافية يسهل الحصول عليها وتكون ذات علاقة بالمتغير المراد تقدير قيمته . مثال ذلك



شكل (١) : شكل الانتشار للعلاقة
بين تكلفة الايجار والبعد عن مركز المدينة

استخدام البيانات الخاصة بالبعد عن مركز المدينة عند دراسة تكلفة الايجار للمناطق المختلفة . ويعطي شكل (١) شكل الانتشار المناظر لبيانات جدول (١) ، حيث يلاحظ وجود اتجاه عام نحو انخفاض تكلفة ايجار المكان كلما بعد عن مركز المدينة . ويكون من الضروري تبعاً لذلك تقدير معدل انخفاض التكلفة نتيجة بعد المكان ، واستخدام ذلك كأساس لتقدير تكلفة ايجار المكان بمعلومية بعده عن مركز المدينة . ويتم ذلك بتقليص شكل الانتشار الى شكل للمتوسطات يتم رسمه بحساب متوسط قيم الظاهرة التابعة (ص) لكل قيمة (أو مجموعة من القيم) للظاهرة المستقلة (س) . ويمثل هذا الشكل العلاقة الرياضية بين قيم س ومتوسط قيم ص المناظرة لها . وتكون هذه العلاقة هي علاقة انحدار ص على س . وقد تم في شكل (٢) حساب متوسط تكلفة الايجار للأماكن التي تبعد مسافة تتراوح بين صفر، ٢٠كم عن مركز المدينة ، وتلك التي تتراوح بين ٢١ ، ٤٠كم وتلك التي تتراوح بين ٤١ ، ٦٠كم . ويمكن الاعتماد على الشكل الناتج للتنبؤ بقيم ص المناظرة لقيم معلومة للمتغير المستقل س. كما يمكن ايضاً تمهيد هذا الشكل وتقريبه بخط مستقيم أو منحنى يمكن التعبير عنه رياضياً بسهولة والاعتماد على ذلك لأغراض التنبؤ .



شكل (٢) : شكل المتوسطات

٢ ـ معنيــان للانحدار

يـرجع استخـدام كلمة انحـدار Regression عند دراسـة العلاقـات بين الـظواهر المختلفـة إلى جالتـون (١٨٢٢ ـ ١٩١١) الـذي لاحظ في مجـال دراساته في علم الوراثة أن الآباء الطوال ينجبـون أبناءاً طـوالاً ، ولكن هؤلاء الأبناء يكونون في المتوسط أقل طولاً من آبـائهم . كذلـك فإن الآبـاء القصار ينجبـون أبناءاً قصـاراً ، ولكن هؤلاء الابناء يكـونون في المتـوسط أطـول من آبائهم . أي يبدو أن هناك اتجاهاً لأن تفقد المجموعات المتميزة في المجتمع هذه الصفة بتعاقب الأجيال، وذلك باقتراب متوسطات هـذه المجموعـات من متوسط المجتمع عموماً . وقد أطلق جالتون على هذه الظاهرة تعبير الانحدار نحو المتوسط .

يمثل شكل المتوسطات المعنى الأول والأساسي لعلاقة انحدار ص على س . وينشأ هذا الشكل بحساب متوسط قيم المتغير التـابع (ص) لكـل قيمة من قيم المتغير المستقل (س). فمثلا ، يعطي جدول (٢) التوزيع التكراري المشترك لمجموعة من ٥٠٠ أسرة حسب عـدد أفراد الأسـرة (س) وما تنفقـه الأسرة شهرياً على الطعام والشراب بمئات الدراهم (ص). إذا أريد تلخيص هذا الجدول بهدف توضيح العلاقة بين س ، ص فإنه يمكن حساب متـوسط لكل عمود من أعمدة الجدول ، أي حساب متوسط قيم (ص) لكـل قيمة من

قيم س . وفي هـذا الصدد ، يـلاحظ أنه عنـدمـا تكـون س = ٢ فـإن ص̄ = ١٤,٥ ، وعندما تكون س = ٣ فإن ص̄ = ١٩,٥ ، وعندما تكون س = ٤ فإن ص̄ = ٢٩,٢ ، وعندما تكون س = ٥ فإن ص̄ = ٣٥,٩ وعندما تكون س = ٦ فـإن ص̄ = ٤٤,٥ . إذا رسمت هـذه القيم للعــلاقـة بين س ، ص̄ ، فــإن المنحنى الناتج يكـون منحنى انحدار ص على س ، ويـظهر ذلـك في شكل (٣) .

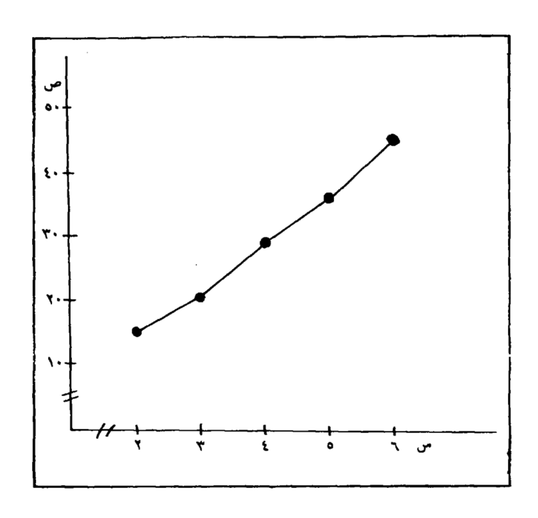
جدول (٢)

التوزيع المشترك للأسر حسب عدد أفراد الأسرة (س)
والمنفق شهرياً على الطعام والشراب (ص)

ص \ س	٢	٣	٤	٥	٦	المجموع
١٠ ـ ١٩	٤٠	٥٠	١٠			١٠٠
٢٠ ـ ٢٩		٥٠	٧٠	٣٠		١٥٠
٣٠ ـ ٣٩			٩٠	٦٠		١٥٠
٤٠ ـ ٤٩				٥٠	٥٠	١٠٠
المجموع	٤٠	١٠٠	١٧٠	١٤٠	٥٠	٥٠٠

وعلى ذلك يمكن القول أن منحنى انحدار ص على س ليس إلا متوسطاً لشكـل الانتشار ، ويجب أن يؤخـذ ذلك في الاعتبـار عند تفسيـر كفاءة هـذا المنحنى في تمثيل العلاقة بين س، ص. ذلك أن منحنى الانحدار ، شأنه في ذلك شأن أي متوسط ، يكون ممثلاً جيداً للبيانات في حالة عدم وجود تشتت واسع للبيانات حوله .

ننتقل الآن إلى المعنى الثاني لعلاقة انحدار ص على س ، وهو المعنى الأكثر شيوعاً في التطبيقات العملية . ويتم الحصول على علاقة الانحدار في هذه الحالة بتوفيق معادلة رياضية تمثل الاتجاه المتوسط في شكل الانتشـار . ولا يدعى أن المعادلة الناتجة ستكون مـطابقة لمنحنى شكـل المتوسـطات ،



شكل (٣) : منحنى انحدار ص على س

وإنما ينظر إليها على أساس أنها تقريب لهذا المنحنى فقط . وتجدر الاشارة إلى أن هذا المعنى الثاني قد أدى إلى انتشار اسلوب الانحدار في التطبيقات العملية المختلفة . ذلك أنه قد يتعذر في كثير من الحالات استخدام المعنى الأول لما يتطلبه ذلك من بيانات كثيرة تستخدم لحساب متوسط قيم ص لكل قيمة من قيم س ، وهي بيانات غالباً ما تكون غير متوافرة .

تختار المعادلة الرياضية التي تمثل علاقة الانحدار بحيث تحتوي على عدد قليل من المعالم بقدر الإمكان وذلك للمحافظة على بساطة أسلوب التحليل . وسوف نكتفي هنا بدراسة الحالة البسيطة التي تمثل فيها هذه المعادلة شكل خط مستقيم . ويجب اعادة التأكيد على أن الهدف الأساسي من انشاء هذه المعادلات هو استخدامها لأغراض التنبؤ بقيم المتغير التابع (ص) .

٣ ـ خط الانحـدار

اذا كان الاتجاه المتوسط للعلاقة بين المتغيرين س ، ص في شكل الانتشار يقترب من علاقة خطية ، فإنه يمكن اتباع أساليب احصائية لتوفيق خط مستقيم مناسب لتمثيل هذه العلاقة . ويقصد بذلك البحث عن خط مستقيم تقع جميع النقط في شكل الانتشار (أو معظمها على الأقل) بالقرب منه . ويتطلب ذلك بالطبع الاتفاق على معيار يمكن الاعتماد عليه كأساس لقياس مدى قرب النقط من الخط .

يعطي شكل (٤) شكل الانتشار المناظر للنقط الخمس التالية :

س :	٥	١٠	١٥	٢٠	٢٥
ص :	١٦	٣٢	٤٤	٤٥	٦٣

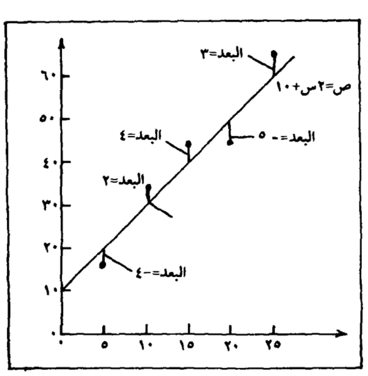
ويعطي الشكل أيضاً خطان هما خط I ، خط II يقترح استخدام أحدهما لتمثيل علاقة الانحدار بين س ، ص . ويلاحظ في هذا الصدد أن خط II أفضل بالتأكيد من خط I نتيجة صغر قيم الاختلافات بين هذا الخط والنقط المختلفة .

شكل (٤)

يمكن أن يقاس مدى قرب نقط شكل الانتشار من الخط المستقيم بمقياس يعتمد على البعد العمودي بين كل نقطة في الشكل وبين الخط المستقيم . ويقصد بهذا البعد العمودي الفرق بين الاحداثي الصادي للنقطة وبين قيمة ص المناظرة لتلك النقطة على الخط المستقيم . فمثلاً يلاحظ أن معادلة خط II في شكل (٤)

هي ص=٢س+١٠ ، وعلى ذلك فإن البعد العمودي بين النقطة (١٠، ٣٢) والخط يتمثل في الفرق بين ٣٢ وبين قيمة ص على الخط عندما تكون س = ١٠ (أي القيمة ٢× ١٠ + ١٠ = ٣٠) وبالتالي يكون البعد العمودي هو ٣٢ - ٣٠ = ٢ . كذلك فإن البعد العمودي بين النقطة (٢٠ ، ٤٥) يكون مساويا للفرق بين ٤٥ وبين قيمة ص على الخط عندما تكون س=٢٠ (أي القيمة ٢×٢٠+١٠=٥٠) وبالتالي يكون البعد العمودي في هذه الحالة هو ٤٥-٥٠= - ٥ . لاحظ أن قيمة البعد الأول موجبة لأن النقطة (١٠، ٣٢) تقع فوق الخط بينما قيمة البعد الثاني سالبة لأن النقطة (٢٠، ٤٥) تقع تحت الخط . ويعطي شكل (٥) قيمة البعد العمودي بين كل نقطة في شكل الانتشار وبين هذا الخط المستقيم .

وتجدر الاشارة إلى أن البعد العمودي بين نقطة ما وبين الخط المستقيم يمثل مقدار الخطأ الذي يحدث عند الاعتماد على الخط للتنبؤ بقيمة ص المناظرة لتلك النقطة . ذلك أن التنبؤ بقيمة ص باستخدام الخط المستقيم يتم بتحديد قيمة ص المناظرة للقيمة المعلومة من قيم س على الخط المستقيم .



شكل (٥) : البعد العمودي لكل نقطة في شكل الانتشار عن الخط المستقيم

ومن المنطقي ، تبعاً لذلك ، أن نحاول الحصول على خط مستقيم تكون أبعاد النقط عنه أصغر ما يمكن . ولما كان من غير الملائم الاعتماد على مجموع هذه الأبعاد (أو متوسطها) كأساس للاختيار بين الخطوط المختلفة نتيجة وجود اشارات موجبة وأخرى سالبة لها ، فإنه يتم تربيع قيم هذه الأبعاد للتخلص من اشاراتها .

ويتخذ مجموع هذه المربعات كمعيار للحكم على مدى جودة تمثيل الخط المستقيم لشكل الانتشار . اذ يكون الخط جيداً كلما كان مجموع مربعات ابعاد النقط عنه صغيراً .

ويمكن أن نحصل رياضياً على أفضل خط لتمثيل شكل الانتشار وفقاً لهذا المعيار ، وهو الخط الذي يكون مجموع مربعات أبعاد النقط عنه أصغر ما يمكن . ويسمى هذا الخط بخط المربعات الصغرى .

مثال (١) : يراد توفيق خط مستقيم لبيانات شكل (٤) ، ومن المطلوب الاختيار بين الخطين ص=٢س+١٠ ، ص=٢,١٤س+٧,٩. يمكن في هذه الحالة الاعتماد على مبدأ المربعات الصغرى بحساب مجموع مربعات ابعاد النقط عن كل خط، ثم اختيار الخط ذو المجموع الأصغر . ويتضح من الحسابات التالية أن الخط ص=٢,١٤س+٧,٩ يكون أفضل تمثيلاً لشكل الانتشار لأن مجموع مربعات ابعاد النقط عنه (٦٥,١٠) أقل من مجموع مربعات ابعاد النقط عن الخط الآخر (٧٠) .

الخط ص = ٢س+١٠				
س	ص	قيمة ص على الخط	البعد العمودي بين النقطة والخط	مربع البعد
٥	١٦	٢٠	٤-	١٦
١٠	٣٢	٣٠	٢	٤
١٥	٤٤	٤٠	٤	١٦
٢٠	٤٥	٥٠	٥-	٢٥
٢٥	٦٣	٦٠	٣	٩
				٧٠

الخط ص=٢,١٤س + ٧,٩		
قيمة ص على الخط	البعد العمودي بين النقطة والخط	مربع البعد
١٨,٦	-٢,٦	٦,٧٦
٢٩,٣	٢,٧	٧,٢٩
٤٠,٠	٤,٠	١٦,٠٠
٥٠,٧	-٥,٧	٣٢,٤٩
٦١,٤	١,٦	٢,٥٦
		٦٥,١٠

ولا يعني ذلك بالطبع أنه من الضروري عند تحديد خط المربعات الصغرى مقارنة مجموع مربعات الأبعاد عن هذا الخط بمجموع المربعات المناظر لكل خط آخر ، لأن ذلك أمر مستحيل عملياً . ذلك أن هناك من النظريات الاحصائية ما يمكن من تحديد هذا الخط مباشرة دون حاجة إلى مقارنته بأي خط آخر .

يمكن اثبات أنه اذا كان أ هو ميل خط المربعات الصغرى وكان ب هو الجزء الذي يقطعه هذا الخط من محور ص ، بحيث أن معادلة الخط هي ص=أس+ب فإنه يمكن الاعتماد على قيم س ، ص المعطاة في البيانات لحساب كل من أ، ب باستخدام الصيغ الرياضية الآتية :

$$أ = \frac{مجـ (س - \bar{س}) (ص - \bar{ص})}{مجـ (س - \bar{س})^{٢}}$$

ب = $\bar{ص}$ - أ $\bar{س}$.

(ويجب التذكير هنا بما سبقت الاشارة إليه من أن مجـ (س - $\bar{س}$) (ص - $\bar{ص}$) يمكن حسابها على الشكل مجـ س ص - ن $\bar{س}$ $\bar{ص}$ ، وأن مجـ (س - $\bar{س}$)^{٢} يمكن حسابها على الشكل (مجـ س^{٢} - ن $\bar{س}^{٢}$) .

مثال (٢) : إذا رجعنا إلى بيانات شكل (٤) ، ولاحظنا أن المتغير (س)

يمثل المسافة التي يقطعها موظف كل صباح في الذهاب إلى عمله (بالكيلومتر) وأن المتغير (ص) يمثل الزمن (بالدقائق) الذي يستغرقه قطع هذه المسافة ويراد استخدام البيانات المعطاة (ن = ٥) لايجاد خط انحدار ص على س بطريقة المربعات الصغرى .

الحل : يلاحظ أن العمليات الحسابية هنا تشبه إلى حد كبير العمليات الحسابية الخاصة بمعامل الارتباط .

يلاحظ أن ن=٥ ، مجـ س=٧٥ ، مجـ ص=٢٠٠ أي أن $\bar{س}$=١٥ ، $\bar{ص}$=٤٠ .

وتظهر بقية الحسابات في الجدول التالي :

س	ص	س - $\bar{س}$	ص - $\bar{ص}$	(س - $\bar{س}$)^{٢}	(س - $\bar{س}$) × (ص - $\bar{ص}$)
٥	١٦	٥ - ١٥ = -١٠	١٦-٤٠ = -٢٤	١٠٠	٢٤٠
١٠	٣٢	١٠-١٥ = -٥	٣٢-٤٠ = -٨	٢٥	٤٠
١٥	٤٤	١٥-١٥ = صفر	٤٤-٤٠ = ٤	صفر	صفر
٢٠	٤٥	٢٠-١٥ = ٥	٤٥-٤٠ = ٥	٢٥	٢٥
٢٥	٦٣	٢٥-١٥= ١٠	٦٣-٤٠ = ٢٣	١٠٠	٢٣٠
٧٥	**٢٠٠**			**٢٥٠**	**٥٣٥**

وعلى ذلك فإن ميل خط الانحدار هو :

$$أ = \frac{مجـ (س - \bar{س}) (ص - \bar{ص})}{مجـ (س - \bar{س})^{٢}} = \frac{٥٣٥}{٢٥٠} = ٢,١٤$$

كما أن الجزء الذي يقطعه خط الانحدار من محور ص هو :

ب = $\bar{ص}$ - أ $\bar{س}$

= ٤٠ - (٢,١٤) (١٥) = ٤٠ - ٣٢,١ = ٧,٩

وتكون معادلة خط انحدار ص على س هي : ص=٢,١٤ س+٧,٩ .

وتجدر الإشارة إلى أن العمليات الحسابية قد تكون أكثر سهولة باستخدام الطريقة البديلة التي تعتمد على مجـ س، مجـ ص، مجـ س ص، مجـ س^{٢} . وذلك على النحو التالي :

س	ص	س^{٢}	س ص
٥	١٦	٢٥	٨٠
١٠	٣٢	١٠٠	٣٢٠
١٥	٤٤	٢٢٥	٦٦٠
٢٠	٤٥	٤٠٠	٩٠٠
٢٥	٦٣	٦٢٥	١٥٧٥
٧٥	**٢٠٠**	**١٣٧٥**	**٣٥٣٥**

حيث يلاحظ أن :

مجـ (س - $\overline{س}$) (ص - $\overline{ص}$) = مجـ س ص - ن $\overline{س}$ $\overline{ص}$

= ٣٥٣٥ - ٥ (١٥) (٤٠) = ٥٣٥

كما أن :

مجـ ($\overline{س}$- س)^{٢} = مجـ س^{٢} - ن $\overline{س}^{٢}$ = ١٣٧٥ - ٥ (١٥)^{٢} = ٢٥٠

وهي نفس النتائج التي حصلنا عليها سابقاً .

ويلاحظ عند تفسير معنى النتائج السابقة أن ميل الخط المستقيم يمثل قيمة التغير في ص المناظرة للتغير في س بمقدار وحدة واحدة . وعلى ذلك فإن القيمة ٢,١٤ التي تمثل ميل خط الانحدار تعني أن زيادة المسافة بمقدار كيلومتر واحد يترتب عليها زيادة في الزمن اللازم لقطعها بمقدار ٢,١٤ دقيقة في المتوسط . ويمكن تبعاً لذلك استخدام المعادلة الناتجة للتنبؤ بقيمة ص المناظرة لقيمة معلومة للمسافة س . فمثلاً عندما تكون المسافة س=١٧,٥كم فإن الزمن اللازم لقطعها يقدر بالقيمة ٢,١٤×١٧,٥ + ٧,٩٠ = ٣٧,٤٥ +

٧,٩٠=٤٥,٣٥دقيقة . كذلك فإن تقـدير الـزمن اللازم لقـطع مسافـة قدرهـا ١٥كم يبلغ ٢,١٤×١٥ + ٧,٩٠=٤٠دقيقة . ويلاحظ في هذا الصدد أن قيمة الزمن المشاهد لمسافة قدرها ١٥ كم يبلغ ٤٤ دقيقة في البيانات المعطاة ، مما يعني أن هناك خطأ في التنبؤ يبلغ في هذه الحالة ٤٤ ـ ٤٠=٤ دقائق .

يجب التنبيه إلى أن معادلة انحدار ص على س هي معادلة تقريبية تصف شكل الانتشار . وعلى ذلك ، فهي معادلة تكون صالحة فقط داخل مدى قيم س المعطاة في البيانات . ويجب توخي الحذر عند الاعتماد عليها للتنبؤ بقيم ص المناظرة لقيم س التي تقع خارج هذا المدى . فإذا أريد في المثال السابق التنبؤ بقيمة ص المناظرة لقيمة س=٤٠كم مثلاً ، فإن استخـدام الخط يعطي القيمة ٢,١٤×٤٠ + ٧,٩=٩٣,٥ دقيقة ولكن الاعتماد على الخط يفترض أن معادلته المحسوبة على أساس المدى (٥ ـ ٢٥) لقيم س تكون أيضاً صـالحة خارج هذا المدى ، وهو افتراض لا يمكن التأكد من سلامته في غياب بيانات اضافية منـاسبة . ويتضـح عدم مـلاءمة معـادلة الانحـدار لتمثيل العـلاقة بين المتغيرين خارج المدى المعطى لقيم س اذا ما أريد التنبؤ بقيمة ص المناظرة لقيمة س=صفر في نفس المثال . إذ تكون قيمة ص في هذه الحـالة مسـاوية ٧,٩٠ دقيقة وهي نتيجة لا معنى لها ، فلا يعقل أن يستغرق قطع مسافة طولها (صفر) كم زمناً موجباً . وعلى ذلك ينبغي كقاعدة عامة عدم استعمال معادلـة الانحـدار خـارج المـدى الـذي حسبت على أسـاسـه إلا اذا كـان هنـاك من المؤشرات ما يدل على صحة المعادلة خارج هذا المدى .

مثال (٣) : بالرجوع إلى البيانات الخاصة بتكلفة الايجار (ص) والبعد عن مركز المدينة (س) المعطاة في جدول (١) ، يراد إيجاد معادلة خط انحدار ص على س، ثم استخدام المعادلة الناتجـة للتنبؤ بتكلفة ايجـار مكان يبعـد مسافة ٤٠كم عن مركز المدينة .

الحل : سوف نعتمد في العمليات الحسابية على مجـ س، مجـ ص، مجـ س ص، مجـ س^{٢} وذلك بسـهولة حساب هذه المقادير .

س	ص	$س^{2}$	س ص
٤	٩٧	١٦	٣٨٨
١٠	٣٥	١٠٠	٣٥٠
١٧	٨٨	٢٨٩	١٤٩٦
١٩	٧٠	٣٦١	١٣٣٠
٣١	٦٣	٩٦١	١٩٥٣
٣٤	٣٦	١١٥٦	١٢٢٤
٣٦	٤٢	١٢٩٦	١٥١٢
٤١	٦٢	١٦٨١	٢٥٤٢
٤٩	٣٥	٢٤٠١	١٧١٥
٥٨	٤١	٣٣٦٤	٢٣٧٨
٢٩٩	**٥٦٩**	**١١٦٢٥**	**١٤٨٨٨**

وعلى ذلك فإن

مجـ (س - $\overline{س}$) (ص - $\overline{ص}$) = مجـ س ص - ن $\overline{س}$ $\overline{ص}$ = ١٤٨٨٨ - ١٠ (٢٩,٩) (٥٦,٩) = -٢١٢٥,١

كما أن

مجـ (س - $\overline{س}$)^{٢} = مجـ س^{٢} - ن $\overline{س}^{٢}$ = ١١٦٢٥ - ١٠ (٢٩,٩)^{٢} = ٢٦٨٤,٩

وبالتالي يكون ميل الخط المستقيم هو :

أ = $\frac{\text{مجـ (س - }\overline{س}\text{) (ص - }\overline{ص}\text{)}}{\text{مجـ (س - }\overline{س}\text{)}^{٢}}$ = - $\frac{٢١٢٥,١}{٢٦٨٤,٩}$ = - ٠,٧٩

كذلك تكون قيمة الجزء الذي يقطعه الخط من محور ص هي :

ب = $\overline{ص}$ - أ $\overline{س}$ = ٥٦,٩ - (-٠,٧٩) (٢٩,٩)

= ٨٠,٥٢

أي أن معادلة انحدار ص على س تأخذ الشكل :

ص = -٧٩, ٠س + ٥٢, ٨٠

إذا أريد التنبؤ بقيمة ص المناظرة لقيمة س=٤٠كم، فإن التنبؤ المطلوب باستخدام خط الانحدار يكون -٧٩, ٠×٤٠ + ٥٢, ٨٠ = ٤٨,٩٢ درهماً .

يجب توخي الحذر عند تقريب النتائج في العمليات الحسابية الوسيطة اللازمة لايجاد معادلة الانحدار . ذلك أن هذا التقريب قد يؤدي إلى تغيير ملموس في النتائج النهائية . فمثلاً إذا كانت قيمة ص̄=١٥٥٠٠ وقيمة س̄=١٥٠٠ فإن قيمة الثابت ب في معادلة الانحدار تكون ١٥٥٠٠-١٥٠٠أ . اذا كانت قيمة أ=٢٧, ١٠ فإن قيمة ب المناظرة تكون ٩٥ . إذا قربت قيمة أ إلى ٣, ١٠ أو إلى ١٠ فإن قيمة ب تكون ٥٠ أو ٥٠٠ على الترتيب . أي أن معادلة الانحدار في هذه الحالة قد تكون ص=٢٧, ١٠س + ٩٥ أو ص= ٣, ١٠ س+٥٠ أو ص=١٠س+٥٠٠ وهي معادلات تختلف عن بعضها بشكل واضح . ويمكن تجنب حدوث ذلك باتباع قاعدة عامة تتمثل في استخدام جميع الأرقام العشرية المتاحة في جميع الخطوات الوسيطة للعمل الحسابي ، وأن يتم التقريب فقط بعد الحصول على النتائج النهائية .

مثال (٣): لدراسة نمط انخفاض قوة الابصار مع تقدم العمر ، جمعت البيانات التالية عن عمر الشخص (س) وقوة ابصاره (ص) لعينة من ١٦ شخصاً :

س :	٢٠	٢٤	٢٥	٢٩	٣٥	٤١
ص :	٤,٠	٤,٧	٤,٤	٤,٢	٥,٦	٣,٦
س :	٥٠	٥١	٦٤	٦٨	٦٩	٦٩
ص :	٢,٩	٢,٣	٣,١	٢,١	٣,٢	٢,٠
س :	٧٣	٧٦	٨١	٨٣		
ص :	٢,٩	٣,٠	٢,٩	٢,٤		

والمطلوب ايجاد معادلة خط انحدار ص على س ، ثم رسم هذه المعادلة فوق شكل الانتشار .

الحل : يمكن للقارىء أن يتأكد حسابياً من صحة النتائج التالية :

$\bar{س}$ = ٥٣,٦٢٥٠ ، $\bar{ص}$ = ٣,٣٣١٢٥٠ ،

مج (س - $\bar{س}$) (ص - $\bar{ص}$) = - ٢٥٣,٩١٢٥٠، مج (س - $\bar{س}$)^{٢} = ٧٢٣٥,٧٥٠.

وعلى ذلك يكون ميل الخط المستقيم هو أ = $\frac{-٢٥٣,٩١٢٥٠}{٧٢٣٥,٧٥٠}$ = - ,٠٣٥٠٩١

كذلك فإن قيمة ب = ٣,٣٣١٢٥٠ - (-,٠٣٥٠٩١) (٥٣,٦٢٥٠) = ٥,٢١٣٠ . أي أن معادلة الانحدار هي :

ص = - ,٠٣٥٠٩١س + ٥,٢١٣٠ .

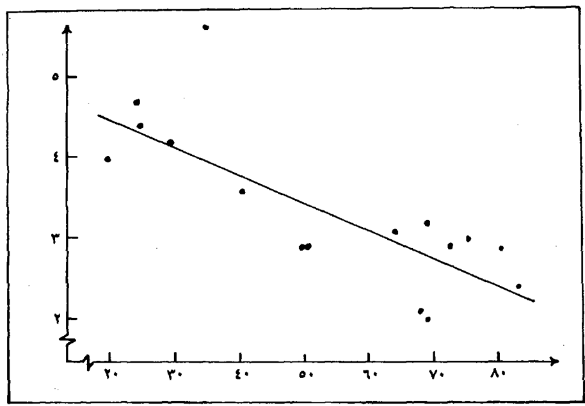
ويمكن للتبسيط تقريب هذه النتيجة النهائية على الشكل :

ص = - ,٠٣٥ س + ٥,٢١ .

(لاحظ عدم اجراء أي تقريب في الخطوات الحسابية الوسيطة) . ولا يؤثر هذا التقريب في كفاءة استخدام المعادلة . فمثلاً إذا أريد التنبؤ بقوة ابصار شخص عمره ٤٠ سنة ، فإن المعادلة الأصلية تعطى - (,٠٣٥٠٩١) (٤٠) + ٥,٢١٣٠ = ٣,٨١ بينما تعطي المعادلة التقريبية القيمة - (,٠٣٥) (٤٠) + ٥,٢١ = ٣,٨١ أيضاً .

يعطي شكل (٦) شكل الانتظار المناظر لهذه البيانات وفوقه خط الانحدار ص = - ,٠٣٥س + ٥,٢١ . وقد تم رسم خط الانحدار بالتعويض عن قيمتين من قيم س داخل المدى المعطى في البيانات وليكن القيم ٢٠ ، ٥٠ مثلاً للحصول على قيم ص المناظرة على الخط . ويلاحظ في هذا الصدد أن

قيمة ص = ٤,٥١ عندما تكون س مساوية ٢٠ كما أن قيمة ص = ٣,٤٦ عندما تكون س = ٥٠ . اذا رسمت النقطتان (٢٠ ، ٤,٥١) ، (٥٠، ٣,٤٦) فإن الخط الموصل بينهما يكون هو خط الانحدار المطلوب .



شكل (٦) : شكل الانتشار وفوقه خط المربعات الصغرى لبيانات مثال (٣)

يلاحظ أن العمليات ا لحسابية المتضمنة في أسلوب الانحدار قد تكون طويلة ومعقدة خاصة إذا كان حجم العينة كبيراً . وتجدر الاشارة إلى أن الحاسبات الآلية تلعب دوراً هاماً في هذا الصدد ، اذ لا تخلو أي مجموعة من برامج الحاسب الآلي الاحصائية من برنامج لحساب معادلات الانحدار . بل يمكن القول أن وجود هذه البرامج قد أدى إلى انتشار استخدام أسلوب الانحدار بشكل واسع في التطبيقات العملية المختلفة .

٤ ـ العلاقة بين خط الانحدار ومعامل الارتباط

هناك علاقة رياضية بين ميل خط انحدار المربعات الصغرى ومعامل الارتباط الخطي . إذ يمكن اثبات أن ميل الخط أ يمكن كتابته على الشكل :

$$أ = ر \times \frac{ع_{ص}}{ع_{س}}$$

حيث ر هو معامل الارتباط ، $ع_{ص}$ هو الانحراف المعياري لقيم المتغير ص، $ع_{س}$ هو الانحراف المعياري لقيم المتغير س. كذلك يمكن إثبات أن معادلة خط الانحدار يمكن أن تكتب على الصورة :

$$ص - \overline{ص} = ر \times \frac{ع_{ص}}{ع_{س}} (س - \overline{س})$$

وتفيد هذه الصيغ في ملاحظة ما يلي :

(أ) أنه عندما تكون قيمة س مساوية $\overline{س}$ فإن قيمة ص المناظرة على خط الانحدار تكون مساوية $\overline{ص}$. ويعني ذلك أن خط الانحدار يمر دائماً بنقطة المتوسطات ($\overline{س}$ ، $\overline{ص}$) .

(ب) اذا كانت قيمة معامل الارتباط ر تساوي الواحد فإن معادلة الخط تصبح $ص - \overline{ص} = \frac{ع_{ص}}{ع_{س}} (س - \overline{س})$، وهي نفس معادلة خط الارتباط التام الذي ناقشناه في الباب السابق. ويتميز هذا الخط بأنه إذا كانت قيمة س تبعد عن متوسطها بعدد من الانحرافات المعيارية، فإن قيمة ص المناظرة تبعد عن متوسطها بنفس عدد الانحرافات المعيارية .

(جـ) قد تكون العلاقة بين س، ص غير تامة. افترض مثلاً أن قيمة معامل الارتباط ر تساوي ٠,٥ في هذه الحالة تكون معادلة خط الانحدار على الشكل :

$$ص - \overline{ص} = ٠,٥ \frac{ع_{ص}}{ع_{س}} (س - \overline{س})$$

اذا كانت قيمة س تساوي $\overline{س} + ع_{س}$، أي أن قيمة س تزيد عن متوسطها بمقدار انحراف معياري واحد ، فإن قيمة ص المناظرة تكون مساوية $\overline{ص} + ٠,٥ ع_{ص}$. كذلك اذا كانت قيمة س تساوي $\overline{س} + ٢ع_{س}$، أي تزيد عن متوسطها بمقدار انحرافين معياريين ، فإن قيمة ص المناظرة تكون مساوية $\overline{ص} + ع_{ص}$، أي تزيد عن متوسطها بمقدار انحراف معياري واحد .

وبصفة عامة يلاحظ أنه إذا كانت قيمة س تبعد عن متوسطها بعدد من الانحرافات المعيارية ، فإن قيمة ص المناظرة تبعد عن متوسطها بنفس العدد من الانحرافات المعيارية مضروباً في قيمة معامل الارتباط ر . ويعني ذلك أنه عند استخدام خط الانحدار لأغراض التنبؤ ، تكون قيمة ص أقرب نسبياً من متوسطها بالمقارنة بقيمة س . وقد سبقت الاشارة إلى هذه الظاهرة عند الحديث عن أصل كلمة انحدار التي نشأت من ملاحظات جالتون في مجال دراساته للعلاقة بين طول الأب وطول الابن . اذ لاحظ جالتون أن طول الابن المقدر من خط الانحدار يبعد عن متوسط هذه الأطوال بمقدار يقل عن بعد طول أبيه عن متوسط الأطوال .

وتسمى هذه الظاهرة بظاهرة الانحدار ولا تعني أن تعاقب الأجيال سيؤدي في النهاية الى تساوي اطوال الاشخاص . وانما يجب أن يكون واضحاً أنها تنشأ نتيجة أسلوب حساب معادلة خط الانحدار ، اذ يكون ميل هذا الخط دائماً أقل من ميل خط الارتباط التام مما يؤدي إلى جعل قيم ص المقدرة من هذا الخط أكثر قرباً نسبياً من متوسطها بالمقارنة بقرب قيم س من متوسطها .

(د) أن معادلة انحدار ص على س تختلف عن معادلة انحدار س على ص.

ذلك أن ميل معادلة انحدار س على ص يكون مساوياً ر × $\frac{ع_س}{ع_ص}$ ويجب أن نتذكر أن خط انحدار ص على س يعمل على تقليل أخطاء التنبؤ إذا كانت ص هي الظاهرة التابعة ، بينما يعمل خط انحدار س على ص على تقليل أخطاء التنبؤ اذا كانت س هي الظاهرة التابعة . وعلى ذلك لا ينبغي استخدام خط انحدار ص على س للتنبؤ بقيمة س ، كما لا ينبغي استخدام خط انحدار س على ص للتنبؤ بقيمة ص .

٥ ـ قياس خطأ التنبؤ عند استخدام معادلة الانحدار

يهتم أسلوب الانحدار بالحصول على علاقات يمكن أن تستخدم للتنبؤ بقيمة المتغير التابع ص . وعلى ذلك يكون من الضروري قياس مدى دقة هذه التنبؤات وذلك بهدف تحديد مدى جودة هذا الأسلوب .

سبقت الاشـارة الى ان خط الانحدار يمثـل متوسط شكـل الانتشـار . وعلى ذلك فان هذا الخط ، شأنه في ذلك شأن المتوسـطات عمومـاً، يكون جيداً كلما كان تشتت النقط حولـه محدوداً . ومن ثم يكـون من المنطقي أن نحسب مقياساً لتشتت نقط شكـل الانتشار حـول الخط ، وذلك للحكم على مدى كفاءة هذا الخط في وصف البيانات .

من الطبيعي في هذه الحالة ان نعتمد على أبعاد النقط المختلفة عن خط الانحدار وذلك بهدف حساب مقياس مشابـه للانحـراف المعياري . ويسمى الفـرق العمودي بين كـل نقطة والخط المستقيم بـالباقي ، ويجب ان يكـون مجمـوع هذه البـواقي مسـاويـاً للصفـر (لأنهـا انحـرافـات قيم عن وسـطهـا الحسابي) . اذا حسب الانحراف المعيـاري لهذه الفـروق فإن النـاتج يمثـل تقـديراً لمتـوسط حجم الخطأ الـذي ينشأ عنـد استخدام المعـادلـة لأغـراض التنبؤ .

مثال ٤ : فيما يلي بيانات عن عمر السيارة بالسنوات (س) وعن سعر بيعها بآلاف الـدراهم (ص) لمجموعـة من عشرة سيـارات بيعت مؤخراً في سوق الحراج .

س :	٦	٥	٤	٥	٣	٢	٣	٤	٥	٣
ص :	٤	٦	٧	٥	١٠	١٢	١١	٩	٧	٩

أوجـد معـادلـة خط انحـدار ص على س ثم احسب قيمـة الانحـراف المعياري لأخطاء التنبؤ بسعر السيارة عند استخدام هذه المعادلة .

الحل : يمكن للقارىء ان يتأكد من صحة الحسابات الآتية :

مجـ س = ٤٠ ، مجـ ص = ٨٠ ، مجـ س^{٢} = ١٧٤ ، مجـ س ص = ٢٩٢

كـذلك فـإن ميل الخط المستقيم أ يسـاوي ـ٢ ، كما أن الجـزء الذي يقطعه هذا الخط من المحور الصادي يساوي ١٦ . أي أن معادلة خط انحدار ص على س هي ص = ـ ٢ س + ١٦ .

لحساب قيمة الانحراف المعياري لأخطاء التنبؤ ، نبدأ بحساب البواقي كما في الجدول التالي :

س	ص	قيمة ص المقدرة من الخط	البواقي
٦	٤	ـ٢ ×٦ +١٦ = ٤	٤ ـ٤ = صفر
٥	٦	ـ٢ ×٥ +١٦ = ٦	٦ ـ٦ = صفر
٤	٧	ـ٢ ×٤ +١٦ = ٨	٧ ـ٨ = ـ١
٥	٥	ـ٢ ×٥ +١٦ = ٦	٥ ـ٦ = ـ١
٣	١٠	ـ٢ ×٣ +١٦ = ١٠	١٠ ـ١٠ = صفر
٢	١٢	ـ٢ ×٢ +١٦ = ١٢	١٢ ـ١٢ = صفر
٣	١١	ـ٢ ×٣ +١٦ = ١٠	١١ ـ١٠ = ١
٤	٩	ـ٢ ×٤ +١٦ = ٨	٩ ـ٨ = ١
٥	٧	ـ٢ ×٥ +١٦ = ٦	٧ ـ٦ = ١
٣	٩	ـ٢ ×٣ +١٦ = ١٠	٩ ـ١٠ = ـ١

ويتم بعد ذلك حساب الانحراف المعياري لقيم البواقي بالطريقة المعتادة لنحصل على :

$$\sqrt{\frac{١}{١٠ - ١}[٦]} = \sqrt{\frac{٦}{٩}} = ٠و٨٢$$

(يلاحظ ان القسمة على (ن ـ ١) عند ايجاد الانحراف المعياري في هذه الحالة ليست دقيقة تماماً لأن المقام يجب ان يختلف قليلاً عن (ن ـ ١) . ولكننا سنتغاضى عن ذلك للتبسيط)

ويستفاد من هذا الانحراف المعياري لوصف تشتت وتوزيع النقط حول خط الانحدار . إذ يمكن مثلاً تطبيق قاعدة تشيبيتشيف والقاعدة العملية ، التي سبقت الاشارة اليها عند دراسة مقاييس التشتت ، وذلك اذا كانت الشروط اللازمة لتطبيق هذه القواعد متوافرة . فمثلاً ، اذا طبقت القاعدة العملية ، فان

حوالي ٦٨٪ من نقط شكل الانتشار تقع على بعد يقل عن انحراف معياري واحد من خط الانحدار ، كذلك فان حوالي ٩٥٪ من نقط شكل الانتشار تبعد عن الخط بأقل من انحرافين معيارين ، وهكذا .

ولما كان أسلوب الانحدار يتضمن الاعتماد على بيانات اضافية (تتمثل في قيم المتغير المستقل س) للتنبوء بقيم المتغير التابع ص ، فإنه من المهم عند دراسة كفاءة هذا الأسلوب تحديد ما اذا كان استخدام هذه البيانات الاضافية قد أدى الى تحسين جودة عملية التنبوء . ويتم ذلك بالتنبوء بقيم ص بافتراض عدم توافر معلومات عن المتغير س ، ثم حساب قيمة الانحراف المعياري للبواقي في هذه الحالة ومقارنة ذلك بقيمة الانحراف المعياري للبواقي عند استخدام معادلة الانحدار .

سبقت الاشارة الى انه في حالة عدم توافر معلومات اضافية ، فانه يكون من المنطقي التنبوء بقيم المتغير التابع ص بالاعتماد على متوسط قيمة ص . وتكون قيم البواقي في هذه الحالة هي القيم ص ـ $\bar{ص}$ ، كما أن الانحراف المعياري لهذه القيم يكون مساوياً للانحراف المعياري لقيم ص أي $ع_{ص}$. (وذلك لأن مجموع البواقي يساوي الصفر) . ويوضح الجدول التالي كيفية حساب البواقي في حالة بيانات مثال (٤) .

س	ص	قيمة ص المقدرة ($\bar{ص}$)	البواقي
٦	٤	٨	٤ ـ ٨ = ـ ٤
٥	٦	٨	٦ ـ ٨ = ـ ٢
٤	٧	٨	٧ ـ ٨ = ـ ١
٥	٥	٨	٥ ـ ٨ = ـ ٣
٣	١٠	٨	١٠ ـ ٨ = ٢
٢	١٢	٨	١٢ ـ ٨ = ٤
٣	١١	٨	١١ ـ ٨ = ٣
٤	٩	٨	٩ ـ ٨ = ١
٥	٧	٨	٧ ـ ٨ = ـ ١
٣	٩	٨	٩ ـ ٨ = ١
			صفر

ويكون الانحراف المعياري لهذه البواقي هو $ع_ص$ ويساوي :

$$ع_ص = \sqrt{\frac{1}{9} [62]} = 2,62$$

حيث يتضح من هذه النتائج ان الاعتماد على عمر السيارة عند التنبوء بسعر بيعها قد أدى الى تخفيض متوسط حجم خطأ التنبوء من ٢,٦٢ الى ٠,٨٢ ، مما يعني أن أسلوب الانحدار قد أدى بالفعل الى تحسين عملية التنبوء .

ومن المنطقي القول بأن أسلوب الانحدار يكون فعالاً في تحسين عملية التنبوء كلما كان الارتباط الخطي بين س ، ص قوياً . ذلك أن نقط شكل الانتشار تتركز في هذه الحالة حول الخط المستقيم وتنخفض بالتالي درجة التشتت ، ومن ثم تقل قيمة الانحراف المعياري لخطأ التنبوء . وفي هذا الصدد ، يمكن ان نثبت رياضياً أن :

$$\text{الانحراف المعياري لخطأ التنبوء} = \sqrt{1 - ر^2} \times ع_ص \ .$$

ويعني ذلك أن المقدار $\sqrt{1 - ر^2}$ يمثل نسبة تقدير متوسط خطأ التنبوء باستخدام معادلة الانحدار الى تقدير متوسط خطأ التنبوء عند اهمال المعادلة . ويلاحظ انه كلما كانت هذه النسبة صغيرة (أي كلما كانت قيمة معامل الارتباط قريبة من الواحد) كلما دل ذلك على جودة خط الانحدار . وبصفة خاصة ، يلاحظ انه إذا كانت قيمة ر = ± ١ فإن الانحراف المعياري لخطأ التنبوء يساوي صفراً ، وتكون معادلة الانحدار قد نجحت نجاحاً تاماً في تمثيل العلاقة بين س ، ص . أما إذا كانت قيمة ر = صفر فإن الانحراف المعياري لخطأ التنبوء يكون مساوياً $ع_ص$ ، مما يعني أن استخدام أسلوب الانحدار في هذه الحالة لا يؤدي الى تحسين عملية التنبوء .

وتجدر الاشارة الى أن قيمة معامل الارتباط بين عمر السيارة وسعر بيعها في البيانات السابقة تساوي ٠,٩٥ ، وبالتالي يمكن حساب قيمة الانحراف المعياري لأخطاء التنبوء مباشرة باستخدام الصيغة الرياضية على الشكل :

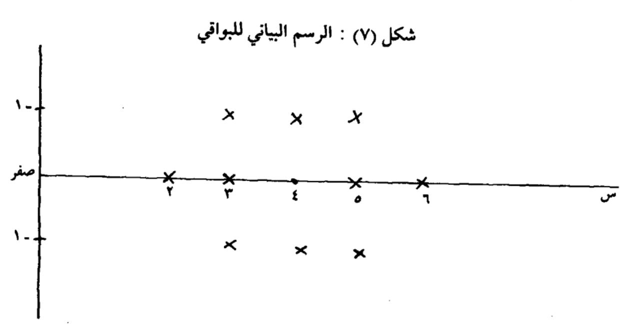
الانحراف المعياري لخطأ التنبوء = $\sqrt{١ - (٠,٩٥)^٢} \times ٢,٦٢$

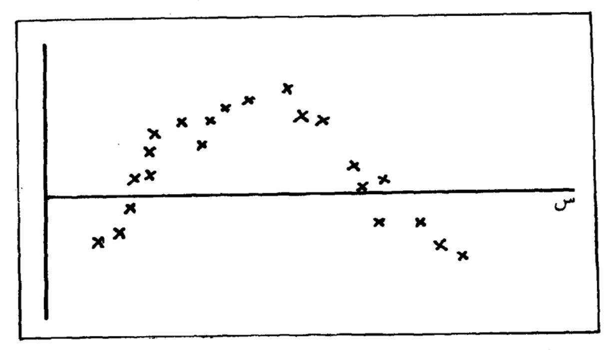
= ٠,٨٢

وهي نفس النتيجة التي حصلنا عليها فيما سبق .

يمكن أيضاً الحكم على جودة خط الانحدار في تمثيل البيانات برسم البواقي بيانياً . وفي هذا الصدد ، ترسم قيم الظاهرة المستقلة س على المحور الأفقي بينما تمثل البواقي المناظرة على المحور الرأسي . يعطي شكل (٧) البواقي الناتجة لبيانات عمر السيارة وسعر بيعها في مثال (٤) . ويلاحظ في هذا الشكل وجود تماثل بين القيم الموجبة والقيم السالبة للبواقي ، كذلك تتوزع هذه القيم بشكل مواز للمحور الأفقي دون ان يكون هناك اتجاه في هذه القيم نحو الصعود أو الهبوط . ويدل ذلك على ان خط الانحدار ممثل جيد للبيانات . ذلك ان خط الانحدار الجيد يؤدي الى بواقي قريبة من خط الصفر ، وموزعة بانتظام على جانبيه ، دون ان يكون هناك اتجاه للصعود في جانب أو للهبوط في جانب آخر . ومن ناحية اخرى يعطي شكل (٨) مثالاً لبواقي تدل على عدم جودة خط الانحدار في تمثيل البيانات ، حيث يلاحظ وجود اتجاه قوي في البواقي نحو الصعود والهبوط . ويدل ذلك على ان خط الانحدار لا يمثل الشكل المناسب لوصف شكل الانتشار وأنه قد يكون من الملائم توفيق منحنى لهذه البيانات بدلاً من خط .

شكل (٧) : الرسم البياني للبواقي





شكل (٨) : مثال لبواقي تدل على عدم جودة خط الانحدار في تمثيل البيانات

٦ ـ ملاحظات عامة

ننتقل الآن الى بعض الملاحظات العامة التي تتعلق بالشروط التي يجب توافرها حتى يمكن استخدام الأساليب التحليلية التي نوقشت في الفصول السابقة ، وهي الشروط التي تضمن ارتفاع كفاءة أسلوب الانحدار في وصف العلاقة بين متغيرين .

(أ) اذا كانت العلاقة بين المتغيرين في شكل الانتشار غير خطية ، فان توفيق خط انحدار لهذه البيانات يؤدي بالتأكيد الى نتائج مضللة . وعلى ذلك ينبغي على الدارسين فحص شكل الانتشار والتأكد من وجود علاقة خطية قبل اجراء الحسابات اللازمة لخط الانحدار .

(ب) يلاحظ ان قيمة الانحراف المعياري لخطأ التنبوء تمثل تقديراً لمتوسط حجم الخطأ الذي ينشأ عند استخدام معادلة الانحدار للتنبوء بقيم ص المناظرة لقيم س المختلفة . ويتضمن الاعتماد على هذا المقياس افتراض عدم وجود اختلافات كبيرة في قيم أخطاء التنبوء عند قيم س المختلفة ، ذلك ان وجود اختلافات كبيرة يعني عدم صلاحية المتوسط لوصف الموضع .

لهذا السبب ، يفترض كشرط أساسي لاستخدام أسلوب الانحدار ان

حجم خطأ التنبوء لا يعتمد على قيمة س . ويمكن التأكد عملياً من تحقق هذا الشرط بدراسة شكل الانتشار وملاحظة ان نمط توزيع النقط حول الخط لا يختلف باختلاف قيمة س .

(جـ) لما كان خط الانحدار يمثل متوسط شكل الانتشار ، وكان الانحراف المعياري لخطأ التنبوء يمثل مقياس التشتت حول هذا المتوسط ، فإنه يجب ان تتوافر الشروط اللازمة في البيانات حتى تكون هذه المقاييس ملائمة لوصف شكل الاختلاف في البيانات . ويقصد بذلك ان يكون توزيع نقط شكل الانتشار حول الخط قريب بقدر الامكان من شكل التوزيع الطبيعي ، وبصفة خاصة يجب ان يكون هذا التوزيع قريب من التماثل وله قمة واحدة .

(د) يرتبط بما سبق أيضاً تأثير القيم الشاذة أو المتطرفة على نتائج التحليل . اذ من المعلوم ان هذه القيم تؤثر بشكل سيء في الوسط الحسابي والانحراف المعياري . وقد سبقت الاشارة الى كيفية معالجة مثل هذه النقط . ونكتفي هنا بالقول بأنه يمكن استبعاد هذه القيم عند اجراء تحليل الانحدار ، مع اعداد تقرير مستقل يصف هذه القيم ودلالاتها .

(هـ) يعتمد أسلوب الانحدار على استخدام قيمة المتغير س للتنبوء بقيمة المتغير التابع ص . ويتضمن ذلك افتراض وجود استقلال بين قيم ص المختلفة ، اذ لا تستخدم احدى قيم ص للتنبوء بقيمة أخرى . وتجدر الاشارة الى أن هذا الفرض قد لا يتحقق اذا كان هناك ترتيب للبيانات . فمثلاً ، عند دراسة درجات الحرارة أثناء النهار ويراد التنبوء بدرجة الحرارة الساعة الواحدة ظهراً ، فإنه لا يكفي الاعتماد على الزمن للتنبوء وانما يمكن أيضاً الاعتماد على درجات الحرارة قبل الساعة الواحدة . في هذه الحالة لا يوجد استقلال بين قيم ص المختلفة ، ولا يمكن تبعاً لذلك استخدام الأساليب التي نوقشت في هذا الباب .

(و) تجدر الاشارة الى أن وجود خط انحدار جيد للمتغير ص على

المتغير س لا يعني بالضرورة وجود علاقة منطقية بين س ، ص . ذلك ان العلاقة الخطية القوية بين س ، ص قد تكون راجعة الى تأثير متغير ثالث يرتبط بكل منهما . فمثلاً ، لاحظ أحد الباحثين وجود علاقة خطية قوية بين مساحة المستطيل وطول محيطه ، وقام تبعاً لذلك بتوفيق خط انحدار لتقدير مساحة المستطيل بمعلومية طول المحيط . ولا شك ان ذلك امر خاطىء لأنه لا توجد علاقة خطية منطقية بين مساحة المستطيل وطول محيطه لأن المساحة مقاسة بوحدات مربعة بينما يقاس المحيط بوحدات عادية . ويمكن ارجاع سبب العلاقة الخطية القوية بينهما الى اعتماد كل منهما على طول وعرض المستطيل . وعلى ذلك ينبغي على الباحثين التعرف على طبيعة العلاقة بين المتغيرين قبل البدء في توفيق خط انحدار لوصف هذه العلاقة .

(ز) يرتبط بما سبق أن علاقة انحدار ص على س لا تعني بالضرورة وجود علاقة سببية بين س ، ص . اذ يجب للتأكد من وجود علاقة سببية ان يكون هناك أساس منطقي لهذه العلاقة ، وان تصمم عملية جمع البيانات لاثبات ذلك باستبعاد تأثير أية عوامل خارجية قد تكون سبباً لوجود علاقة الانحدار ، وذلك بالاعتماد على الأساليب الاحصائية لتصميم التجارب وتحليل نتائجها .

(ك) ناقشنا في هذا الباب معادلة انحدار متغير تابع ص على متغير مستقل واحد س . وتجدر الاشارة الى انه يمكن تعميم هذه الأساليب لايجاد معادلة انحدار ص على اكثر من متغير مستقل واحد . ويكون الهدف من اضافة متغيرات مستقلة اخرى الى المعادلة هو تحسين قدرة المعادلة على التنبوء بقيم ص . وتسمى المعادلة الناتجة في هذه الحالة بمعادلة الانحدار المتعدد ، ولن يتسع المجال هنا لمناقشة هذا النوع من معادلات الانحدار .

تمرينات

١ ـ اذا كان هناك البيانات التالية عن س ، ص :

س :	ـ١	صفر	١
ص :	ـ ١	ـ ١	٢

(أ) احسب قيمة معامل الارتباط بين س ، ص .

(ب) أوجد معادلة خط انحدار ص على س .

(جـ) ارسم شكل الانتشار للبيانات ، وارسم فوقه خط الانحـدار الذي حصلت عليـه في (ب) . تأكـد من أن خط الانحـدار يمـر بنقـطة المتوسطات ($\bar{س}$ ، $\bar{ص}$) .

٢ ـ قام أحد الأشخاص بتسجيل كمية الوقود التي يضعها في سيارته والمسافة التي تقطعها السيارة باستخدام هذه الكمية في كل مرة يملأ فيها سيارتـه بالوقود خلال العام الماضي ، فكانت البيانات التالية :

كمية الوقود باللترات : وسطها = ٣٣,٥ وانحرافها المعياري = ٢ .

المسافة بالكيلومترات : وسطها = ٤٥٠ وانحرافها المعياري = ٣٠

وكانت قيمة معامل الارتباط ر = ٠,٤

(أ) أوجد معادلة انحدار كمية الوقود على المسافة .

(ب) احسب قيمة الانحراف المعياري لخطأ التنبؤ لهذه المعادلة .

(جـ) هل تعتقد أن معادلة الانحدار معادلة جيدة ؟ اشرح سبب الاجابة .

٣ ـ فيما يلي بيانات عن متغيرين س ، ص

س :	٠,٧	١,٠	٠,٧	٠,٠
ص :	٠,٧	٠,٠	ـ ٠,٦	ـ ١,٠

(أ) احسب معادلة خط انحدار ص على س .

(ب) ارسم شكل الانتشار لهـذه البيانـات ، ثم ارسم فوقـه الخط الذي تحصل عليه في (أ) . علق على الشكل الناتج .

٤ ـ يعتمد سكان إحدى جزر المحيط الهندي في غذائهم بصفة أساسية على نوع من الأسماك يتميز بارتفاع نسبة تركز الزئبق فيه . ويترتب على ذلك ارتفاع مستويات تركز الزئبق في أجسام هؤلاء السكان . يرغب أحد الباحثين في دراسة ما إذا كان ارتفاع مستويات تركز الزئبق لدى الأمهات في هذه الجزيرة ينتقل إلى الجنين خلال فترة الحمل . قام هذا الباحث بقياس درجة تركز الزئبق لدى الأم ولدى مولودها وذلك لعينة من ١٠ أمهات فكانت البيانات التالية :

	مستوى تركز الزئبق	
رقم	لدى الأم	لدى المولود
١	١٤	١٢
٢	٢٧	٤٨
٣	٧	١١
٤	٨	٦٢
٥	٦	٦
٦	١٧	١٨
٧	١٥	١٩
٨	١٦	٢١
٩	٧	٧
١٠	١٧	١٤

استخدم أسلوب الانحدار للاجابة على تساؤلات هذا الباحث .

٥ ـ في دراسة عن العلاقة بين درجة الحرارة (س) وزمن حفظ نوع معين من الطعام عند هذه الدرجة قبل فساده (ص) وجد أن :

ن = ١٥ ، مجـ س = ١٤٢٥ ، مجـ ص = ١٠,٦٨

مجـ س ص = ٩٨٧,٦٥ ، مجـ $س^{٢}$ = ١٣٩٠٣٧,٢٥

(أ) أوجد معادلة خط انحدار ص على س .

(ب) ما هو قيمة التغير المتوقع في الزمن اذا زادت درجة الحرارة بمقدار عشر درجات ؟

٦ ـ فيما يلي بيانات عن المنفق على الدعاية والاعلان (س) وعن قيمة المبيعات (ص) لستة أنواع من المشروبات الغازية وذلك خلال عام معين :

س (بآلاف الدراهم) :	٥٠١	٣٢١	١٣٧	٣٨٧	١٣٦	٣٠٣
ص (بملايين الدراهم):	٣,٤١	٢,٠٥	,٩٦	٢,٨٩	,٧٤	٢,٢١

(أ) ارسم شكل الانتشار لهذه البيانات .

(ب) احسب قيمة معامل الارتباط بين س ، ص .

(جـ) أوجد معادلة خط انحدار ص على س .

(د) احسب القيمة المتوقعة لمبيعات أحد المشروبات الغازية اذا بلغ المنفق على الدعاية لهذا المشروب خلال العام ١٥٠ ألف درهم .

٧ ـ (أ) « تكون اشارة ميل خط الانحدار مشابهة لاشارة معامل الارتباط المحسوب من نفس البيانات » . اشرح سبب ذلك .

(ب) « من الخطورة الاعتماد على خط الانحدار للتنبؤ بقيمة ص المناظرة لقيمة من قيم س تكون اكبر بكثير أو أقل بكثير عن قيم س المعطاة في البيانات » . اشرح سبب ذلك .

٨ ـ قام أحد المديرين في شركة استهلاكية باختيار عينة عشوائية من ١٠ بائعين لدراسة العلاقة بين عدد سنوات خبرة البائع (س) وبين قيمة مبيعاته السنوية (ص) . وقد اتضح من شكل الانتشار ان هناك علاقة خطية بين س ، ص .

(أ) افترض ان معامل الارتباط بين س ، ص في العينة هو ر = ٠,٧٥ وأن متوسط المبيعات السنوية للعمال = ١٠٠ ، وأن الانحراف المعياري لهذه المبيعات = ٨ . اذا كانت خبرة أحد العاملين تقع

فوق المتوسط بمقدار انحرافين معيارين ، فما هي قيمة التنبؤ بالمبيعات السنوية له .

(ب) اذا كانت خبرة أحد البائعين تقع تحت المتوسط بمقدار ١,٥ انحرافاً معيارياً وقدر تبعاً لذلك ان قيمة مبيعاته السنوية تقع تحت المتوسط بمقدار انحراف معياري واحد ، فما هي قيمة معامل الارتباط في هذه الحالة ؟

٩ ـ فيما يلي بيانات عن السمك (س) وقوة الشد (ص) لخمسة أنواع من السلوك الحديدية :

س (بالمليمتر):	٢٠٠	٢١٠	٢٢٠	٢٣٠	٢٤٠
ص (بالرطل):	٨١٣,٧	٧٨٥,٣	٩٦٠,٤	١١١٨,٠	١٠٧٦,٢

(أ) اذا علم ان مجـ (س ـ $\overline{س}$)^{٢} = ١٠٠٠ ، مجـ (س ـ $\overline{س}$) (ص ـ $\overline{ص}$) = ٨٥٧٧ ، فأوجد معادلة انحدار ص على س .

(ب) اذا علم أن : ١ رطل =٠,٤٥٣٦ كجم، وتم تحويل قيم ص إلى كيلوجرامات ، فما هي معادلة خط الانحدار في هذه الحالة .

١٠ ـ في دراسة عن العلاقة بين درجة تلوث مياه البحيرات ومعدلات وفاة أسماك هذه البحيرات جمعت بيانات عن درجة التلوث (س) وعن نسبة الأسماك الباقية على قيد الحياة لمدة معلومة (ص) في ٩ بحيرات ، فكانت البيانات التالية :

س :	١٠	١٠	٢٠	٢٠	٢٥	٢٧	٢٧	٣١	٥٠
ص :	٨٥	٩٢	٨٥	٩٦	٨٧	٨٠	٩٠	٥٩	٦٢

وجد أن خط انحدار ص على س هو ص = ـ ٠,٧٨ س + ١٠٠,٧٩ .

(أ) ارسم شكل المتوسطات لهذه البيانات ، وارسم فوقه خط الانحدار . علق على مدى ملاءمة خط الانحدار لوصف شكل المتوسطات .

(ب) احسب قيمة الانحراف المعياري لخطأ التنبؤ .

(جـ) ارسم شكلاً بيانياً للبواقي وعلق على الشكل الناتج .

(د) اذا تم تقريب معادلة الخط على الشكل ص = ـ س + ١٠٠ ، هل تكون قيمة الانحراف المعياري لخطأ التنبؤ في هذه الحالة اكبر أم أقل من القيمة المحسوبة في (ب) ؟ وضح سبب اجابتك .

١١ ـ اذا كان معلوماً ان الوسط الحسابي لدرجة الطالب في مبادىء الاحصاء في امتحان نصف الفصل = ٦٠ وأن الانحراف المعياري = ١٥ . كذلك فإن الوسط الحسابي لدرجة الطالب في الامتحان النهائي والانحراف المعياري لهذه الدرجات يساوي ٦٠، ١٥ على الترتيب أيضاً . وكانت قيمة معامل الارتباط بين درجة الطالب في الامتحانين تساوي ٠,٥ .

(أ) يراد التنبؤ بدرجة أحد الطلبة في الامتحان النهائي في الحالات التالية :

(i) اذا علم ان درجته في امتحان نصف الفصل تساوي ٦٠

(ii) اذا علم ان درجته في امتحان نصف الفصل تساوي ٧٥

(iii) اذا علم ان درجته في امتحان نصف الفصل تساوي ٣٠

(iv) اذا كانت درجة الطالب في امتحان نصف الفصل غير معلومة .

(ب) احسب قيمة الانحراف المعياري لخطأ التنبؤ عند استخدام معادلة الانحدار في هذه الحالة .

ـ ١٢ ـ (أ) « يكون ميل خط الانحدار دائماً أقل من ميل خط الارتباط التام » هل تتفق مع صحة هذه العبارة ؟ ولماذا ؟

(ب) يرغب مدير مكتب القبول بالجامعة في اختيار أسلوب للتنبؤ

بدرجات الطلبة المقبولين بعد عام من التحاقهم بالجامعة . لديه اسلوبان أحدهما يتعرض لأخطاء انحرافها المعياري = ١٢ بينما يتعرض الآخر لأخطاء انحرافها المعياري = ٧ . أي الأسلوبين يكون اكثر ملاءمة ؟ ولماذا ؟

١٣ ـ وجد في دراسة ما أن معامل الارتباط بين المستوى التعليمي للزوج والمستوى التعليمي للزوجة يساوي ٠,٥٠ وأن متوسط عدد سنوات التعليم للزوج والزوجة = ١٢ وأن الانحراف المعياري لعدد سنوات تعليم الزوج والانحراف المعياري لسنوات تعليم الزوجة يساوي كل منهما ٣ سنوات .

(أ) قدر عدد سنوات التعليم لامرأة يكون عدد سنوات تعليم زوجها مساو ١٨ سنة .

(ب) قدر عدد سنوات التعليم لرجل يكون عدد سنوات تعليم زوجته مساو ١٥ سنة .

(جـ) يبدو من هذه النتائج ان الرجال الأكثر تعليماً يتزوجون نساءاً على مستوى تعليمي أقل . وفي نفس الوقت فان هؤلاء النساء يتزوجون رجالاً على مستوى تعليمي اكثر انخفاضاً . هل تتفق مع هذا التفسير للنتائج ؟ اشرح سبب اجابتك .

١٤ ـ تقوم الجامعة بعقد امتحان لتحديد المستوى في اللغة الانجليزية للطلبة المقبولين الجدد . في دراسة للعلاقة بين درجة الطالب في هذا الامتحان ودرجته في اللغة الانجليزية في امتحان الثانوية العامة وجد أن :

امتحان تحديد المستوى : المتوسط = ٦٥ والانحراف المعياري = ٨

امتحان الثانوية العامة : المتوسط = ٦٧ والانحراف المعياري = ١٠

وكانت قيمة معامل الارتباط = ٠,٦٠

(أ) اذا أريد تخمين درجة أحد الطلبة في امتحان تحديد المستوى ، وذلك دون توافر أية معلومات أخرى عن هذا الطالب ، فما هو هذا التخمين ؟

(ب) ما هي قيمة الانحراف المعياري لخطأ التنبوء في هذه الحالة ؟

(جـ) كرر (أ) ، (ب) اذا علم ان درجة الطالب في امتحان الثانوية العامة تساوي ٧٠ .

١٥ ـ في دراسة عن العلاقة بين طول الشخص عند العمر ٦ سنوات وطوله عند العمر ١٨ سنة وجد أن :

الطول عند العمر ٦ سنوات : الوسط الحسابي = ٤٦ بوصةوالانحراف المعياري = ١,٧ بوصة .

الطول عند العمر ١٨ سنة : الوسط الحسابي = ٧٠ بوصةوالانحراف المعياري = ٢,٥ بوصة .

وكانت قيمة معامل الارتباط = ٠,٨٠

(أ) أوجد معادلة انحدار الطول عند ١٨ سنة على الطول عند ٦ سنوات ، ثم احسب قيمة الانحراف المعياري لخطأ التنبوء في هذه المعادلة .

(ب) اوجد معادلة انحدار الطول عند ٦ سنوات على الطول عند ١٨ سنة ، ثم احسب قيمة الانحراف المعياري لخطأ التنبؤ فى هذه المعادلة .

(جـ) قدر طول شخص عند العمر ١٨ سنة اذا علم أن طوله عند العمر ٦ سنوات يساوي ٥٠ بوصة .

(د) قدر طول شخص عند العمر ٦ سنوات اذا علم ان طوله عند العمر ١٨ سنة يساوي ٧٥ بوصة .

١٦ ـ فيما يلي بيانات عن معدل الوفاة بسبب حوادث السيارات في بلد ما خلال الفترة (١٩٧٢ ـ ١٩٨٠) :

السنة	١٩٧٢	١٩٧٣	١٩٧٤	١٩٧٥	١٩٧٦	١٩٧٧	١٩٧٨	١٩٧٩	١٩٨٠
معدل الوفاة في الألف	١,٨	٢,٠	٢,١	٢,٢	٢,٤	٢,٥	٢,٧	٢,٧	٢,٧

(أ) اوجد معادلة خط انحدار معدل الوفاة على الزمن .

(ب) ارسم شكل البواقي في هذه الحالة وعلق عليه .

(جـ) احسب قيمة الانحراف المعياري لخطأ التبنوء .

(د) استخدام معادلة الانحدار للتنبوء بمعدل الوفاة في هذا البلد في عام ١٩٨٥ وفي عام ٢٠٠٠ . ما هي الفروض اللازمة لصحة هذه التنبؤات ؟ وهل تعتقد ان هذه الفروض متحققة ؟ وضح سبب اجابتك .

١٧ ـ فيما يلي بيانات عن نصيب الفرد في تكلفة الخدمات الصحية في دولة ما في السنوات (٧٠ ـ ١٩٧٩) :

السنة :	١٩٧٠	١٩٧١	١٩٧٢	١٩٧٣	١٩٧٤	١٩٧٥	١٩٧٦	١٩٧٧	١٩٧٨	١٩٧٩
نصيب الفرد بالدراهم :	٢٩٥	٣٢٥	٣٥٠	٤٠٠	٤٥٠	٥٥٠	٦٠٠	٦٧٥	٧٢٥	٨٠٠

(أ) أوجد معادلة خط انحدار نصيب الفرد على الزمن.

(ب) ارسم شكل الانتشار لهذه البيانات ، وارسم فوقه خط الانحدار وعلق على الشكل الناتج .

(جـ) هل تصلح المعادلة الناتجة للتنبوء بنصيب الفرد في تكلفة الخدمات الصحية في عام ٢٠٠٠ ؟ وضح سبب اجابتك .

مراجع مختارة

1. Anderson, D. R and D. J. Sweeny (1984), *Statistics for Business and Economics, West.*
2. Besag, F. and P.L. Besag (1985), *Statistics for the Helping Professionsn, Sage.*
3. Brase, C. and C. Brase (1983), *Understandal lc Statistics: Concepts and Methods*, Heath.
4. Chase, C.I.(1984) *Elementary Statistical Procedures*, 3rd., Mc Graw Hill.
5. Christensen, H.B (1977), *Statistics: Step by Step*, Haughton Mifflin.
6. Daniel, W. (1983), *Biostatistics: A Foundation For Analysis in the Health Sciences*, Wiley.
7. Daniel, W. and J. C. Terrell (1982), *Business Statistics: Basic Concepts and Methodology*, 3rd ed., Houghton Mifflin.
8. Ehrenberg, A. S. C. (1975) *Data Reduction*, Wiley.
9. Freedman, D., R. Pisani and R. Purves (1978), *Statistics, W.W. Nordon.*
10. Hays, W.L. (1981), *Statistics* , Holt, Rinehart and Wilson.
11. Johson, R. and G. K. Bhattachayya (1985), *Statistics: Principles and Methods, Wiley.*
12. Mendenhall, W. (1983), *Introduction to Protability and Statistics,* 6th ed, Duxbury Press.
13. Neter, J. and W. Wasserman (1982), *Applied Statistics*, 2nd ed., Irwin.
14. Rustagi, J.S. (1985), *Introduction to Statistical Analysis for Business Decisions* , Irwin.
16. Walpole, R.E. (1982), *Introduction to Statistics*, 3éd ed., Macmil lan.
17. Wonnacott, T. and R. J. Wonnacott (1977), *Introduction Statistics* , 3rd ed., Wiley.

طبع في دار النفائس ص.ب ٦٣٤٧-١١ هاتف ٨١٠١٩٤ - بيروت

تطلب منشوراتنا من :

دار القلم دبي — طريق النفق - بناية الشيخ راشد القديمة - برقيا : قلمكم
هاتف : ٤٣٣٨٨٦ - ص.ب : ١١٨١٧ - تلكس : ٤٦٩٦١ - حتى أي أم

دار القلم الكويت — شارع الثور - بجوار وزارة الخارجية - عمارة الشوف
ص.ب : ٢٠١٤٦ - هـ : ٢٤٥٨٤٧٨ - ٢٤٥٧٤٠٧ - برقيا : توزيعكم

